arXiv:0910.0352v2 [math-ph] 14 Oct 2009

Quantum Physics, Relativity, and Complex Spacetime: Towards a New Synthesis

Gerald Kaiser
Department of Mathematics
University of Massachusetts–Lowell
kaiser@wavelets.com • http://wavelets.com

ABSTRACT

The positivity of the energy in relativistic quantum mechanics implies that wave functions can be continued analytically to the forward tube $\mathcal{T}_+$ in complex spacetime. For Klein-Gordon particles, we interpret $\mathcal{T}_+$ as an extended (8D) classical phase space containing all 6D classical phase spaces as symplectic submanifolds. The evaluation maps $e_z : f \to f(z)$ of wave functions on $\mathcal{T}_+$ are relativistic coherent states reducing to the Gaussian coherent states in the nonrelativistic limit. It is known that no covariant probability interpretation exists for Klein-Gordon particles in real spacetime because the time component of the conserved "probability current" can attain negative values even for positive-energy solutions. We show that this problem is solved very naturally in complex spacetime, where $|f(x-iy)|^2$ is interpreted as a probability density on all 6D phase spaces in $\mathcal{T}_+$ which, when integrated over the "momentum" variables y, gives a conserved spacetime probability current whose time component is a positive regularization of the usual one. Similar results are obtained for Dirac particles, where the evaluation maps e_z are spinor-valued relativistic coherent states. For free quantized Klein-Gordon and Dirac fields, the above formalism extends to n-particle/antiparticle coherent states whose scalar products are Wightman functions. The 2-point function plays the role of a reproducing kernel for the one-particle and antiparticle subspaces.

Originally published as a book in 1990 by North-Holland, Amsterdam
© Gerald Kaiser 2003. All rights reserved.

Unified field Theory

In the beginning there was Aristotle
And objects at rest tended to remain at rest
And objects in motion tended to come to rest
And soon everything was at rest
And God saw that it was boring.

Then God created Newton
And objects at rest tended to remain at rest
But objects in motion tended to remain in motion
And energy was conserved and momentum was conserved
and matter was conserved
And God saw that it was conservative.

Then God created Einstein
And everything was relative
And fast things became short
And straight things became curved
And the universe was filled with inertial frames
And God saw that it was relatively general
but some of it was especially relative.

Then God created Bohr
And there was the $\mathcal{P}$rinciple
And the principle was $\mathcal{Q}$uantum
And all things were quantized
But some things were still $\mathcal{R}$elative
And God saw that it was confusing.

Then God was going to create $\mathcal{M}$
And $\mathcal{M}$ would have unified
And $\mathcal{M}$ would have fielded a theory
And all would have been one
But it was the seventh day
And God rested
And objects at rest tend to remain at rest.

Adapted from a poem by Tim Joseph

CONTENTS

PREFACE

The idea of complex spacetime as a unification of spacetime and classical phase space, suitable as a possible geometric basis for the synthesis of Relativity and quantum theory, first occured to me in 1966 while I was a physics graduate student at the University of Wisconsin. In 1971, during a seminar I gave at Carleton University in Canada, it was pointed out to me that the formalism I was developing was related to the coherent–state representation, which was then unknown to me. This turned out to be a fortunate circumstance, since many of the subsequent developments have been inspired by ideas related to coherent states. My main interest at that time was to formulate relativistic coherent states.

In 1974, I was struck by the appearance of tube domains in axiomatic quantum field theory. These domains result from the analytic continuation of certain functions (vacuum expectaion values) associated with the theory to complex spacetime, and powerful methods from the theory of several complex variables are then used to prove important properties of these functions in *real* spacetime. However, the complexified spacetime itself is usually not regarded as having any physical significance. What intrigued me was the possibility that these tube domains may, in fact, have a direct physical interpretation as (extended) classical phase spaces. If so, this would give the idea of complex spacetime a firm physical foundation, since in quantum field theory the complexification is based on solid physical principles. It could also show the way to the construction of relativistic coherent states. These ideas were successfully worked out in 1975-76, culminating in a mathematics thesis in 1977 at the University of Toronto entitled "Phase–Space Approach to Relativistic Quantum Mechanics."

Up to that point, the theory could only describe free particles. The next goal was to see how interactions could be added. Some progress in this direction was made in 1979-80, when a natural way was found to extend gauge theory to complex spacetime. Further progress came during my sabbatical in 1985-86, when a method was developed for extending quantized fields themselves (rather than their vacuum expectation values) to complex spacetime. These ideas have so far produced no "hard" results, but I believe that they are on the right path.

Although much work remains to be done, it seems to me that enough structure is now in place to justify writing a book. I hope that this volume will be of interest to researchers in theoretical and

mathematical physics, mathematicians interested in the structure of fundamental physical theories and assorted graduate students searching for new directions. Although the topics are fairly advanced, much effort has gone into making the book self–contained and the subject matter accessible to someone with an understanding of the rudiments of quantum mechanics and functional analysis.

A novel feature of this book, from the point of view of mathematical physics, is the special attention given to "signal analysis" concepts, especially time–frequency localization and the new idea of *wavelets*. It turns out that relativistic coherent states are similar to wavelets, since they undergo a Lorentz contraction in the direction of motion. I have learned that engineers struggle with many of the same problems as physicists, and that the interplay between ideas from quantum mechanics and signal analysis can be very helpful to both camps. For that reason, this book may also be of interest to engineers and engineering students.

The contents of the book are as follows. In chapter 1 the simplest examples of coherent states and time–frequency localization are introduced, including the original "canonical" coherent states, windowed Fourier transforms and wavelet transforms. A generalized notion of frames is defined which includes the usual (discrete) one as well as continuous resolutions of unity, and the related concept of reproducing kernels is discussed.

In chapter 2 a new, algebraic approach to orthonormal bases of wavelets is formulated. An operational calculus is developed which simplifies the formalism considerably and provides insights into its symmetries. This is used to find a *complex structure* which explains the symmetry between the low– and the high–frequency filters in wavelet theory. In the usual formulation, this symmetry is clearly evident but appears to be accidental. Using this structure, complex wavelet decompositions are considered which are analogous to analytic coherent–state representations.

In chapter 3 the concept of generalized coherent states based on Lie groups and their homogeneous spaces is reviewed. Considerable attention is given to holomorphic (analytic) coherent–state representations, which result from the possibility of Lie group complexification. The rotation group provides a simple yet non–trivial proving ground for these ideas, and the resulting construction is known as the "spin coherent states." It is then shown that the group associated with the Harmonic oscillator is a weak contraction limit (as the spin $s \to \infty$) of the rotation group and, correspondingly, the canonical co-

herent states are limits of the spin coherent states. This explains why the canonical coherent states transform naturally under the dynamics generated by the harmonic oscillator.

In chapter 4, the interactions between phase space, quantum mechanics and Relativity are studied. The main ideas of the phase–space approach to relativistic quantum mechanics are developed for free particles, based on the relativistic coherent–state representations developed in my thesis. It is shown that such representations admit a covariant probabilistic interpretation, a feature absent in the usual spacetime theories. In the non–relativistic limit, the representations are seen to "contract" smoothly to representations of the Galilean group which are closely related to the canonical coherent–state representation. The Gaussian weight functions in the latter are seen to emerge from the geometry of the mass hyperboloid.

In chapter 5, the formalism is extended to quantized fields. The basic tool for this is the *Analytic–Signal transform,* which can be applied to an arbitrary function on $\mathbb{R}^n$ to give a function on $\mathbb{C}^n$ which, although not in general analytic, is "analyticity–friendly" in a certain sense. It is shown that even the most general fields satisfying the Wightman axioms generate a complexification of spacetime which may be interpreted as an extended classical phase space for certain special states associated with the theory. Coherent–state representations are developed for free Klein–Gordon and Dirac fields, extending the results of chapter 4. The analytic Wightman two–point functions play the role of reproducing kernels. Complex–spacetime densities of observables such as the energy, momentum, angular momentum and charge current are seen to be *regularizations* of their counterparts in real spacetime. In particular, Dirac particles do not undergo their usual Zitterbewegung. The extension to complex spacetime separates, or *polarizes,* the positive– and negative–frequency parts of free fields, so that Wick ordering becomes unnecessary. A functional–integral representation is developed for quantized fields which combines the coherent–state representations for particles (based on a finite number of degrees of freedom) with that for fields (based on an infinite number of degrees of freedom).

In chapter 6 we give a brief account of some ongoing work, beginning with a review of the idea of holomorphic gauge theory. Whereas in real spacetime it is not possible to derive gauge potentials and gauge fields from a (fiber) metric, we show how this can be done in complex spacetime. Consequently, the analogy between General Relativity and gauge theory becomes much closer in complex spacetime than it is in real spacetime. In the "holomorphic" gauge class, the

relation between the (non–abelian) Yang–Mills field and its potential becomes linear due to the cancellation of the non–linear part which follows from an integrability condition. Finally, we come full circle by generalizing the Analytic–Signal transform and pointing out that this generalization is a higher–dimensional version of the wavelet transform which is, moreover, closely related to various classical transforms such as the Hilbert, Fourier–Laplace and Radon transforms.

I am deeply grateful to G. Emch for his continued help and encouragement over the past ten years, and to John Klauder and Ray Streater for having read the manuscript carefully and made many invaluable comments, suggestions and corrections. (Any remaining errors are, of course, entirely my responsibility.) I also thank D. Buchholtz, F. Doria, D. Finch, S. Helgason, I. Kupka, Y. Makovoz, J. E. Marsden, M. O'Carroll, L. Rosen, M. B. Ruskai and R. Schor for miscellaneous important assistance and moral support at various times. Finally, I am indebted to L. Nachbin, who first invited me to write this volume in 1981 (when I was not prepared to do so) and again in 1985 (when I was), and who arranged for a tremendously interesting visit to Brazil in 1982. *Quero também agradecer a todos os meus colegas Brasileiros!*

Suggestions to the Reader

The reader primarily interested in the phase–space approach to relativistic quantum theory may on first reading skip chapters 1–3 and read only chapters 4–6, or even just chapter 4 and either chapters 5 or 6, depending on interest. These chapters form a reasonably self–contained part of the book. Terms defined in the previous chapters, such as "frame," can be either ignored or looked up using the extensive index. The index also serves partially as a glossary of frequently used symbols. The reader primarily interested in signal analysis, time–frequency localization and wavelets, on the other hand, may read chapters 1 and 2 and skip directly to sections 5.2 and 6.2. The mathematical reader unfamiliar with the ideas of quantum mechanics is urged to begin by reading section 1.1, where some basic notions are developed, including the Dirac notation used throughout the book.

Chapter 1
COHERENT–STATE REPRESENTATIONS

1.1. Preliminaries

In this section we establish some notation and conventions which will be followed in the rest of the book. We also give a little background on the main concepts and formalism of non–relativistic and relativistic quantum mechanics, which should make this book accessible to non–specialists.

1. Spacetime and its Dual

In this book we deal almost exclusively with flat spacetime, though we usually let space be $\mathbb{R}^s$ instead of $\mathbb{R}^3$, so that spacetime becomes $X = \mathbb{R}^{s+1}$. The reason for this extension is, first of all, that it involves little cost since most of the ideas to be explored here readily generalize to $\mathbb{R}^{s+1}$, and furthermore, that it may be useful later. Many models in constructive quantum field theory are based on two– or three–dimensional spacetime, and many currently popular attempts to unify physics, such as string theories and Kaluza–Klein theories, involve spacetimes of higher dimensionality than four or (on the string world–sheet) two–dimensional spacetimes. An event $x \in X$ has coordinates

$$x = (x^\mu) = (x^0, x^j), \tag{1}$$

where $x^0 \equiv t$ is the time coordinate and x^j are the space coordinates. Greek indices run from 0 to s, while latin indices run from 1 to s. If we think of x as a *translation vector,* then X is the vector space of all translations in spacetime. Its *dual* X^* is the set of all linear maps $k\colon X \to \mathbb{R}$. By linearity, the action of k on x (which we denote by kx instead of $k(x)$) can be written as

$$kx = \sum_{\mu=0}^{s} k_\mu x^\mu \equiv k_\mu x^\mu, \tag{2}$$

where we adopt the Einstein summation convention of automatically summing over repeated indices. Usually there is no relation between

x and k other than the pairing $(x,k) \mapsto kx$. But suppose we are given a scalar product on X,

$$x \cdot x' = g_{\mu\nu} x^\mu x'^\nu \tag{3}$$

where $(g_{\mu\nu})$ is a non–degenerate matrix. Then each x in X defines a linear map $x^*\colon X \to \mathbb{R}$ by $x^*(x') = x \cdot x'$, thus giving a map $*\colon X \to X^*$, with

$$(x^*)_\nu \equiv x_\nu = g_{\mu\nu} x^\mu. \tag{4}$$

Since $g_{\mu\nu}$ is non–degenerate, it also defines a scalar product on X^*, whose metric tensor is denoted by $g^{\mu\nu}$. The map $x \to x^*$ establishes an isomorphism between the two spaces, which we use to identify them. If x denotes a set of inertial coordinates in free spacetime, then the scalar product is given by

$$g_{\mu\nu} = \mathrm{diag}(c^2, -1, -1, \cdots, -1)$$

where c is the speed of light. X, together with this scalar product, is called *Minkowskian* or *Lorentzian* spacetime.

It is often convenient to work in a single space rather than the dual pair X and X^*. Boldface letters will denote the spatial parts of vectors in X^*. Thus $x = (t, -\mathbf{x})$, $k = (k_0, \mathbf{k})$ and

$$x \cdot x' = c^2 t t' - \mathbf{x} \cdot \mathbf{x}' \quad \text{and} \quad kx = k \cdot x^* = k_0 t - \mathbf{k} \cdot \mathbf{x}, \tag{5}$$

where $\mathbf{x} \cdot \mathbf{x}'$ and $\mathbf{k} \cdot \mathbf{x}$ denote the usual Euclidean inner products in $\mathbb{R}^s$.

2. Fourier Transforms

The Fourier transform of a function $f\colon X \to \mathbb{C}$ (which, to avoid analytical subtleties for the present, may be assumed to be a Schwartz test function; see Yosida [1971]) is a function $\hat{f}\colon X^* \to \mathbb{C}$ given by

$$\hat{f}(k) = \int_X dx\, e^{2\pi i k x} f(x) \tag{6}$$

where $dx \equiv dt\, d^s\mathbf{x}$ is Lebesgue measure on X. f can be reconstructed from $\hat{f}$ by the inverse Fourier transform, denoted by $\check{}$ and given by

$$f(x) = \int_{X^*} dk\, e^{-2\pi i k x} \hat{f}(k) \equiv (\hat{f})\check{}(x), \tag{7}$$

where $dk = dk_0\, d^s\mathbf{k}$ denotes Lebesgue measure on $X^* \approx X$. Note that the presence of the 2π factor in the exponent avoids the usual need for factors of $(2\pi)^{-(s+1)/2}$ or $(2\pi)^{-s-1}$ in front of the integrals. Physically, k represents a *wave vector:* $k_0 \equiv \nu$ is a frequency in cycles per unit time, and k_j is a wave number in cycles per unit length. Then the interpretation of the linear map k: $X \to \mathbb{R}$ is that $2\pi kx$ is the total radian phase gained by the plane wave $g(x') = \exp(-2\pi ikx')$ through the spacetime translation x, i.e. $2\pi k$ "measures" the radian phase shift. Now in pre–quantum relativity, it was realized that the *energy* E combines with the *momentum* $\mathbf{p}$ to form a vector $p \equiv (p_\mu) = (E, \mathbf{p})$ in X^*. Perhaps the single most fundamental difference between classical mechanics and quantum mechanics is that in the former, matter is conceived to be made of "dead sets" moving in space while in the latter, its microscopic structure is that of *waves* descibed by complex–valued wave functions which, roughly speaking, represent its distribution in space in probabilistic terms. One important consequence of this difference is that while in classical mechanics one is free to specify position and momentum independently, in quantum mechanics a complete knowledge of the distribution in space, i.e. the wave function, determines the distribution in momentum space via the Fourier transform. The classical energy is re–interpreted as the frequency of the associated wave by Planck's *Ansatz*,

$$p_0 \equiv E = 2\pi\hbar\nu \tag{8}$$

where $\hbar$ is Planck's constant, and the classical momentum is re–interpreted as the wave–number vector of the associated wave by De Broglie's relation,

$$\mathbf{p} = 2\pi\hbar\mathbf{k}. \tag{8'}$$

These two relations are unified in relativistic terms as $p_\mu = 2\pi\hbar k_\mu$. Since a general wave function is a superposition of plane waves, each with its own frequency and wave number, the relation of energy and momentum to the the spacetime structure is very different in quantum mechanics from what is was in classical mechanics: They become operators on the space of wave functions:

$$(P_\mu f)(x) = \int_{X^*} dk\, p_\mu e^{-2\pi ikx} \hat{f}(k) = i\hbar \frac{\partial}{\partial x^\mu} f(x), \tag{9}$$

or, in terms of x^*,

$$P_0 = i\hbar\frac{\partial}{\partial t} \quad \text{and} \quad P_k = -i\hbar\frac{\partial}{\partial x_k}. \tag{9′}$$

This is, of course, the source of the uncertainty principle. In terms of energy–momentum, we obtain the "quantum–mechanical" Fourier transform and its inverse,

$$\begin{aligned} \hat{f}(p) &= \int_X dx\, e^{ipx/\hbar} f(x) \\ f(x) &= (2\pi\hbar)^{-s-1} \int_{X^*} dp\, e^{-ipx/\hbar} \hat{f}(p). \end{aligned} \tag{10}$$

If $f(x)$ satisfies a differential equation, such as the Schrödinger equation or the Klein–Gordon equation, then $\hat{f}(p)$ is supported on an s–dimensional submanifold P of X^* (a paraboloid or two–sheeted hyperboloid, respectively) which can be parametrized by $\mathbf{p} \in \mathbb{R}^s$. We will write the solution as

$$f(x) = (2\pi\hbar)^{-s} \int_P d\mu(p)\, e^{-ipx/\hbar} \hat{f}(p), \tag{11}$$

where $\hat{f}(p)$ is, by a mild abuse of notation, the "restriction" of $\hat{f}$ to P (actually, $|\hat{f}(p)|^2$ is a density on P) and $d\mu(p) \equiv \rho(p)\, d^s\mathbf{p}$ is an appropriate invariant measure on P. For the Schrödinger equation $\rho(p) \equiv 1$, whereas for the Klein–Gordon equation, $\rho(p) = |\, p_0\,|^{-1}$. Setting $t = 0$ then shows that $\hat{f}(p)$ is related to the initial wave function by

$$f(\mathbf{x}, 0) = \big(\rho(p)\hat{f}\big)\check{}(\mathbf{x}), \tag{12}$$

where now "$\check{}$" denotes the the s–dimensional inverse Fourier transform of the function $\hat{f}$ on $P \approx \mathbb{R}^s$.

We will usually work with "natural units," i.e. physical units so chosen that $\hbar = c = 1$. However, when considering the non–relativistic limit ($c \to \infty$) or the classical limit ($\hbar \to 0$), c or $\hbar$ will be re–inserted into the equations.

3. Hilbert Space

Inner products in Hilbert space will be linear in the *second* factor and antilinear in the first factor. Furthermore, we will make some discrete use of Dirac's very elegant and concise *bra–ket* notation, favored by physicists and often detested or misunderstood by mathematicians.

As this book is aimed at a mixed audience, I will now take a few paragraphs to review this notation and, hopefully, convince mathematicians of its correctness and value. When applied to coherent–state representations, as opposed to representations in which the position– or momentum operators are diagonal, it is perfectly rigorous. (The bra–ket notation is problematic when dealing with distributions, such as the generalized eigenvectors of position or momentum, since it tries to take the "inner products" of such distributions.)

Let $\mathcal{H}$ be an arbitrary complex Hilbert space with inner product $\langle \cdot, \cdot \rangle$. Each element $f \in \mathcal{H}$ defines a bounded linear functional $f^*\colon \mathcal{H} \to \mathbb{C}$ by

$$f^*(g) = \langle f, g \rangle. \tag{13}$$

The Riesz representation theorem guarantees that the converse is also true: Each bounded linear functional $L : \mathcal{H} \to \mathbb{C}$ has the form $L = f^*$ for a unique $f \in \mathcal{H}$. Define the *bra* $\langle f|$ corresponding to f by

$$\langle f| \, = f^* : \mathcal{H} \to \mathbb{C}. \tag{14}$$

Similarly, there is a one-to-one correspondence between vectors $g \in \mathcal{H}$ and linear maps

$$|g\rangle\colon \mathbb{C} \to \mathcal{H} \tag{15}$$

defined by

$$|g\rangle(\lambda) = \lambda g, \quad \lambda \in \mathbb{C}, \tag{16}$$

which will be called *kets*. Thus elements of $\mathcal{H}$ will be denoted alternatively by g or by $|g\rangle$. We may now consider the composite map *bra–ket*

$$\langle f | \, g \rangle : \mathbb{C} \to \mathbb{C}, \tag{17}$$

given by

$$\begin{aligned} \langle f \, | \, g \rangle(\lambda) &= f^*(\lambda g) = \lambda f^*(g) \\ &= \lambda \langle f, \, g \rangle. \end{aligned} \tag{18}$$

Therefore the "bra–ket" map is simply the multiplication by the inner product $\langle f, \, g \rangle$ (whence it derives its name). Henceforth we will identify these two and write $\langle f | \, g \rangle$ for both the map and the inner product. The reverse composition

$$|g\rangle\langle f| : \mathcal{H} \to \mathcal{H} \tag{19}$$

may be viewed as acting on kets to produce kets:

$$|g\rangle\langle f|\big(|h\rangle\big) = |g\rangle\big(\langle f|h\rangle\big). \tag{20}$$

To illustrate the utility of this notation, as well as some of its pitfalls, suppose that we have an orthonormal basis $\{g_n\}$ in $\mathcal{H}$. Then the usual expansion of an arbitrary vector f in $\mathcal{H}$ takes the form

$$f = \sum_n g_n \langle\, g_n \,|\, f \,\rangle \equiv \sum_n |\, g_n \,\rangle\langle\, g_n \,|\, f \,\rangle, \tag{21}$$

from which we have the "resolution of unity"

$$\sum_n |\, g_n \,\rangle\langle\, g_n \,| \, = I \tag{22}$$

where I is the identity on $\mathcal{H}$ and the sum converges in the strong operator topology. If $\{h_n\}$ is a second orthonormal basis, the relation between the expansion coefficients in the two bases is

$$\langle\, h_n \,|\, f \,\rangle = \langle\, h_n \,|\, If \,\rangle = \sum_m \langle\, h_n \,|\, g_m \,\rangle\langle\, g_m \,|\, f \,\rangle. \tag{23}$$

In physics, vectors such as g_n are often written as $|\, n \,\rangle$, which can be a source of great confusion for mathematicians. Furthermore, functions in $L^2(\mathbb{R}^s)$, say, are often written as $f(x) = \langle\, x \,|\, f \,\rangle$, with $\langle\, x \,|\, x' \,\rangle = \delta(x - x')$, as though the $|\, x \,\rangle$'s formed an orthonormal basis. This notation is very tempting; for example, the Fourier transform is written as a "change of basis,"

$$\langle\, k \,|\, f \,\rangle = \int dx \, \langle\, k \,|\, x \,\rangle\langle\, x \,|\, f \,\rangle \tag{24}$$

with the "transformation matrix" $\langle\, k \,|\, x \,\rangle = \exp(2\pi i k x)$. One of the advantages of this notation is that it permits one to think of the Hilbert space as "abstract," with $\langle\, g_n \,|\, f \,\rangle, \langle\, h_n \,|\, f \,\rangle, \langle\, x \,|\, f \,\rangle$ and $\langle\, k \,|\, f \,\rangle$ merely different "representations" (or "realizations") of the same vector f. However, even with the help of distribution theory, this use of Dirac notation is unsound, since it attempts to extend the Riesz representation theorem to distributions by allowing inner products of

them. (The "vector" $\langle\, x \,|$ is a distribution which evaluates *test functions* at the point x; as such, $|\, x' \,\rangle$ does not exist within modern–day distribution theory.) We will generally abstain from this use of the bra–ket notation.

Finally, it should be noted that the term "representation" is used in two distinct ways: (a) In the above sense, where abstract Hilbert–space vectors are represented by functions in various function spaces, and (b) in connection with groups, where the action of a group on a Hilbert space is represented by operators.

This notation will be especially useful when discussing frames, of which coherent–state representations are examples.

1.2. Canonical Coherent States

We begin by recalling the original coherent-state representations (Bargmann [1961], Klauder [1960, 1963a, b], Segal [1963a]). Consider a spinless non-relativistic particle in $\mathbb{R}^s$ (or $s/3$ such particles in $\mathbb{R}^3$), whose algebra of observables is generated by the position operators X_k and momentum operators $P_k, k = 1, 2, \ldots s$. These satisfy the "canonical commutation relations"

$$[X_k, X_l] = 0, \qquad [P_k, P_l] = 0, \qquad [X_k, P_l] = i\delta_{kl} I, \tag{1}$$

where I is the identity operator. The operators $-iX_k$, $-iP_k$ and $-iI$ together form a *real* Lie algebra known as the *Heisenberg algebra*, which is irreducibly represented on $L^2(\mathbb{R}^s)$ by

$$X_k f(x) = x_k f(x), \qquad P_k f(x) = -i\frac{\partial}{\partial x_k} f(x), \tag{2}$$

the *Schrödinger representation.*

As a consequence of the above commutation relations between X_k and P_k, the position and momentum of the particle obey the *Heisenberg uncertainty relations,* which can be derived simply as follows. The expected value, upon measurement, of an observable represented by an operator F in the state represented by a wave function $f(x)$ with $\|f\| = 1$ (where $\|\cdot\|$ denotes the norm in $L^2(\mathbb{R}^s)$) is given by

$$\langle\, F \,\rangle = \langle\, f \,|\, Ff \,\rangle. \tag{3}$$

In particular, the expected position– and momentum coordinates of the particle are $\langle X_k \rangle$ and $\langle P_k \rangle$. The uncertainties Δ_{X_k} and Δ_{P_k} in position and momentum are given by the variances

$$\begin{aligned} \Delta^2_{X_k} &= \langle (X_k - \langle X_k \rangle)^2 \rangle = \langle X_k^2 \rangle - \langle X_k \rangle^2 \\ \Delta^2_{P_k} &= \langle (P_k - \langle P_k \rangle)^2 \rangle = \langle P_k^2 \rangle - \langle P_k \rangle^2. \end{aligned} \tag{4}$$

Choose an arbitrary constant b with units of area (square length) and consider the operators

$$A_k = X_k + ibP_k = x_k + b\frac{\partial}{\partial x_k}. \tag{5}$$

Notice that although A_k is non–Hermitian, it is *real* in the Schrödinger representation. Let

$$\bar{z}_k \equiv \langle f \,|\, A_k f \rangle = \langle X_k \rangle + ib\langle P_k \rangle, \tag{6}$$

where $\bar{z}_k$ denotes the complex–conjugate of z_k. Then for $\delta A_k \equiv A_k - \bar{z}_k I$ we have $\langle \delta A_k \rangle = 0$ and

$$0 \leq \|\delta A_k f\|^2 = \Delta^2_{X_k} + b^2\Delta^2_{P_k} - b. \tag{7}$$

The right–hand side is a quadratic in b, hence the inequality for all b demands that the discriminant be non–positive, giving the uncertainty relations

$$\Delta_{X_k}\Delta_{P_k} \geq \frac{1}{2}. \tag{8}$$

Equality is attained if and only if $\delta A_k f = 0$, which shows that the *only* minimum–uncertainty states are given by wave functions $f(x)$ satisfying the eigenvalue equations

$$A_k f = \big(x_k + b\frac{\partial}{x_k}\big) f = \bar{z}_k f \tag{9}$$

for some real number b (which may actually depend on k) and some $z \in \mathbb{C}^s$. But square–integrable solutions exist only for $b > 0$, and then there is a unique solution (up to normalization) χ_z for each $z \in \mathbb{C}^s$. To simplify the notation, we now choose $b = 1$. Then A_k and A_k^* satisfy the commutation relations

$$[A_k, A_l] = 0, \qquad [A_k, A_l^*] = 2\delta_{kl} I, \tag{10}$$

and χ_z is given by

$$\chi_z(x') = N \exp[-\bar{z}^2/4 + \bar{z} \cdot x' - x'^2/2], \tag{11}$$

where the normalization constant is chosen as $N = \pi^{-s/4}$, so that $\|\chi_z\| = 1$ for $z = 0$. Here $\bar{z}^2$ is the (complex) inner product of $\bar{z}$ with itself. Clearly χ_z is in $L^2(\mathbb{R}^s)$, and if $z = x - ip$, then

$$\langle X_k \rangle = x_k \text{ and } \langle P_k \rangle = p_k \tag{12}$$

in the state given by χ_z. The vectors χ_z are known as the *canonical coherent states*. They occur naturally in connection with the harmonic oscillator problem, whose Hamiltonian can be cast in the form

$$H = \frac{1}{2m}\big(P^2 + m^2\omega^2 X^2\big) = \frac{1}{2}m\omega^2 A^* \cdot A + \frac{s\omega}{2} \tag{13}$$

with

$$A_k = X_k + iP_k/m\omega \tag{14}$$

(thus $b = 1/m\omega$). They have the remarkable property that if the initial state is χ_z, then the state at time t is $\chi_{z(t)}$ where $z(t)$ is the orbit in phase space of the corresponding *classical* harmonic oscillator with initial data given by z. These states were discovered by Schrödinger himself [1926], at the dawn of modern quantum mechanics. They were further investigated by Fock [1928] in connection with quantum field theory and by von Neumann [1931] in connection with the quantum measurement problem. Although they span the Hilbert space, they do not form a basis because they possess a high degree of linear dependence, and it is not easy to find complete, linearly independent subsets. For this reason, perhaps, no one seemed to know quite what to do with them until the early 1960's, when it was discovered that what really mattered was not that they form a basis but what we shall call a *generalized frame*. This allows them to be used in generating a *representation* of the Hilbert space by a space of analytic functions, as explained below. The frame property of the coherent states (which will be studied and generalized in the following sections and in chapter 3) was discovered independently at about the same time by Klauder, Bargmann and Segal. Glauber [1963a,b] used these vectors with great effectiveness to extend the concept of optical coherence to the domain of quantum electrodynamics, which was made

necessary by the discovery of the laser. He dubbed them "coherent states," and the name stuck to the point of being generic. (See also Klauder and Sudarshan [1968].) Systems of vectors now called "coherent" may have nothing to do with optical coherence, but there is at least one unifying characteristic, namely their frame property (next section).

The coherent-state representation is now defined as follows: Let $\mathcal{F}$ be the space of all functions

$$\begin{aligned}\tilde{f}(z) \equiv \langle \chi_z \,|\, f\rangle &= \int d^s x' \, \overline{\chi_z(x')} f(x') \\ &= N \int d^s x' \, \exp[-z^2/2 + z\cdot x' - x'^2/2] f(x')\end{aligned} \tag{15}$$

where f runs through $L^2(\mathbb{R}^s)$. Because the exponential decays rapidly in x', $\tilde{f}$ is entire in the variable $z \in \mathbb{C}^s$. Define an inner product on $\mathcal{F}$ by

$$\langle \tilde{f} \,|\, \tilde{g}\rangle_{\mathcal{F}} \equiv \int_{\mathbb{C}^s} d\mu(z) \, \overline{\tilde{f}(z)} \tilde{g}(z), \tag{16}$$

where $z \equiv x - ip$ and

$$d\mu(z) = (2\pi)^{-s} \exp(-\bar{z}\cdot z/2) \, d^s x \, d^s p. \tag{17}$$

Then we have the following theorem relating the inner products in $L^2(\mathbb{R}^s)$ and $\mathcal{F}$.

Theorem 1.1. *Let $f, g \in L^2(\mathbb{R}^s)$ and let $\tilde{f}, \tilde{g}$ be the corresponding entire functions in $\mathcal{F}$. Then*

$$\langle \tilde{f} \,|\, \tilde{g}\rangle_{\mathcal{F}} = \langle f \,|\, g\rangle_{L^2}. \tag{18}$$

Proof. To begin with, assume that f is in the Schwartz space $\mathcal{S}(\mathbb{R}^s)$ of rapidly decreasing smooth test functions. For $z = x - ip$, we have

$$\chi_z(x') = N \exp[-\bar{z}^2/2 + x^2/2 - (x'-x)^2/2 + ip\cdot x'],$$

hence

$$\tilde{f}(x-ip) = N \exp[(x^2+p^2)/4 + ipx/2] \big(\exp[-(x'-x)^2/2] f\big)\hat{}\,(p) \tag{19}$$

where $\hat{}$ denotes the Fourier transform with respect to x'. Thus by Plancherel's theorem (Yosida [1971]),

$$\begin{aligned}(2\pi)^{-s} \int d^s p \, \exp(-p^2/2) |\tilde{f}(x - ip)|^2 \\ = N^2 \exp(x^2/2) \int d^s x' \exp[-(x' - x)^2] \, |f(x')|^2.\end{aligned} \tag{20}$$

Therefore

$$\begin{aligned}\int d\mu(z) \, |\tilde{f}(z) \, |^2 \\ = N^2 \int d^s x \int d^s x' \exp[-(x' - x)^2] \, |f(x') \, |^2 \\ = \int d^s x' \, | \, f(x') \, |^2\end{aligned} \tag{21}$$

after exchanging the order of integration. This proves that

$$\|\tilde{f}\|^2_{\mathcal{F}} = \|f\|^2_{L^2} \tag{22}$$

for $f \in \mathcal{S}(\mathbb{R}^s)$, hence by continuity also for arbitrary $f \in L^2(\mathbb{R}^s)$. By polarization the result can now be extended from the norms to the inner products. ∎

The relation $f \leftrightarrow \tilde{f}$ can be summarized neatly and economically in terms of Dirac's bra-ket notation. Since

$$\overline{\tilde{f}(z)} = \overline{\langle \chi_z | f \rangle} = \langle f | \chi_z \rangle, \tag{23}$$

theorem 1 can be restated as

$$\int_{\mathbb{C}^s} d\mu(z) \, \langle f | \chi_z \rangle \langle \chi_z | g \rangle = \langle f | g \rangle. \tag{24}$$

Dropping the bra $\langle f|$ and ket $|g\rangle$, we have the operator identity

$$\int_{\mathbb{C}^s} d\mu(z) \, |\chi_z\rangle \langle \chi_z| = I, \tag{25}$$

where I is the identity operator on $L^2(\mathbb{R}^s)$ and the integral converges at least in the sense of the *weak operator topology*,* i.e. as a quadratic form. In Klauder's terminology, this is a *continuous resolution of unity*. A general operator B on $L^2(\mathbb{R}^s)$ can now be expressed as an integral operator $\tilde{B}$ on $\mathcal{F}$ as follows:

$$\begin{aligned}(\tilde{B}\tilde{f})(z) &\equiv (Bf)\tilde{}(z)\\ &= \langle\, \chi_z \,|\, Bf \,\rangle\\ &= \langle\, \chi_z \,|\, BIf \,\rangle\\ &= \int_{\mathbb{C}^s} d\mu(w)\, \langle\, \chi_z \,|\, B\chi_w \,\rangle \langle\, \chi_w \,|\, f \,\rangle\\ &\equiv \int_{\mathbb{C}^s} d\mu(w)\, \tilde{B}(z,\bar{w})\tilde{f}(w).\end{aligned} \tag{26}$$

Particularly simple representations are obtained for the basic position– and momentum operators. We get

$$\begin{aligned}X_k\chi_z(x') &\equiv x'_k\chi_z(x')\\ &= \left(\frac{\partial}{\partial\bar{z}_k} + \frac{\bar{z}_k}{2}\right)\chi_z(x')\end{aligned} \tag{27}$$

$$\begin{aligned}P_k\chi_z &= -i(A_k - X_k)\chi_z\\ &= i\left(\frac{\partial}{\partial\bar{z}_k} - \frac{\bar{z}_k}{2}\right)\chi_z,\end{aligned}$$

thus

$$\tilde{X}_k\tilde{f} = \left(\frac{\partial}{\partial z_k} + \frac{z_k}{2}\right)\tilde{f} \tag{28}$$

$$\tilde{P}_k\tilde{f} = -i\left(\frac{\partial}{\partial z_k} - \frac{z_k}{2}\right)\tilde{f}.$$

* As will be shown in a more general context in the next section, under favorable conditions the integral actually converges in the *strong* operator topology.

Hence X_k and P_k can be represented as differential rather than integral operators.

As promised, the continuous resolution of the identity makes it possible to *reconstruct* $f \in L^2(\mathbb{R}^s)$ from its transform $\tilde{f} \in \mathcal{F}$:

$$f = If = \int_{\mathbb{C}^s} d\mu(z) \, | \, \chi_z \, \rangle \langle \, \chi_z \, | \, f \, \rangle, \tag{29}$$

that is,

$$f(x') = \int_{\mathbb{C}^s} d\mu(z) \, \chi_z(x') \tilde{f}(z). \tag{30}$$

Thus in many respects the coherent states behave like a *basis* for $L^2(\mathbb{R}^s)$. But they differ from a basis in at least one important respect: They cannot all be linearly independent, since there are uncountably many of them and $L^2(\mathbb{R}^s)$ (and hence also $\mathcal{F}$) is separable. In particular, the above reconstruction formula can be used to express χ_z in terms of all the χ_w's:

$$\chi_z = \int_{\mathbb{C}^s} d\mu(w) \, | \, \chi_w \, \rangle \langle \, \chi_w \, | \, \chi_z \, \rangle. \tag{31}$$

In fact, since entire functions are determined by their values on some discrete subsets Γ of $\mathbb{C}^s$, we conclude that the corresponding subsets of coherent states $\{\chi_z \, | \, z \in \Gamma\}$ are already complete since for any function f orthogonal to them all, $\tilde{f}(z) = 0$ for all $z \in \Gamma$ and hence $\tilde{f} \equiv 0$, which implies $f = 0$ a.e. For example, if Γ is a regular lattice, a necessary and sufficient condition for completesness is that Γ contain at least one point in each Planck cell (Bargmann et al., [1971]), in the sense that the spacings Δx_k and Δp_k of the lattice coordinates $z_k = x_k + ip_k$ satisfy $\Delta x_k \Delta p_k \leq 2\pi\hbar \equiv 2\pi$. It is no accident that this looks like the uncertainty principle but with the inequality going "the wrong way." The exact coefficient of $\hbar$ is somewhat arbitrary and depends on one's definition of uncertainty; it is possible to define measures of uncertainty other than the standard deviation. (In fact, a preferable—but less tractable—definition of uncertainty uses the notion of *entropy,* which involves all moments rather than just the second moment. See Bialynicki–Birula and Mycielski [1975] and Zakai [1960].) The intuitive explanation is that if $\tilde{f}$ gets "sampled" at least once in every Planck cell, then it is uniquely determined since the uncertainty principle limits the amount of variation which can take place within such a cell. Hence the set of all coherent states

is *overcomplete*. We will see later that reconstruction formulas exist for some discrete subsystems of coherent states, which makes them as useful as the continuum of such states. This ability to synthesize continuous and discrete methods in a single representation, as well as to bridge quantum and classical concepts, is one more aspect of the appeal and mystery of these systems.

1.3. Generalized Frames and Resolutions of Unity

Let M be a set and μ be a measure on M (with an appropriate σ–algebra of measurable subsets) such that $\{M, \mu\}$ is a σ–finite measure space. Let $\mathcal{H}$ be a Hilbert space and $h_m \in \mathcal{H}$ be a family of vectors indexed by $m \in M$.

Definition. The set

$$\mathcal{H}_M \equiv \{h_m \,|\, m \in M\} \tag{1}$$

is a *generalized frame* in $\mathcal{H}$ if

1. the map h: $m \mapsto h_m$ is weakly measurable, i.e. for each $f \in \mathcal{H}$ the function $\tilde{f}(m) \equiv \langle\, h_m \,|\, f \,\rangle$ is measurable, and
2. there exist constants $0 < A \leq B$ such that

$$A\|f\|^2 \leq \int_M d\mu(m)\, |\, \langle\, h_m \,|\, f \,\rangle\, |^{\,2} \leq B\|f\|^2 \qquad \forall f \in \mathcal{H}. \tag{2}$$

$\mathcal{H}_M$ is a *frame* (see Young [1980] and Daubechies [1988a]) in the special case when M is countable and μ is the counting measure on M (i.e., it assigns to each subset of M the number of elements contained in it). In that case, the above condition becomes

$$A\,\|f\|^2 \leq \sum_{m\in M} |\, \langle\, h_m \,|\, f \,\rangle\, |^{\,2} \leq B\,\|f\|^2 \qquad \forall f \in \mathcal{H}. \tag{3}$$

We will henceforth drop the adjective "generalized" and simply speak of "frames." The above case where M is countable will be refered to as a *discrete* frame.

If $A = B$, the frame $\mathcal{H}_M$ is called *tight*. The coherent states of the last section form a tight frame, with $M = \mathbb{C}^s, d\mu(m) = d\mu(z), h_m = \chi_z$ and $A = B = 1$.

Given a frame, let T be the map taking vectors in $\mathcal{H}$ to functions on M defined by

$$(Tf)(m) \equiv \langle\, h_m \,|\, f \,\rangle \equiv \tilde{f}(m), \quad f \in \mathcal{H}. \tag{4}$$

Then the frame condition states that Tf is square-integrable with respect to $d\mu$, so that T defines a map

$$T\colon \mathcal{H} \to L^2(d\mu), \tag{5}$$

with

$$A\,\|f\|^2 \le \|Tf\|^2_{L^2(d\mu)} \le B\,\|f\|^2. \tag{6}$$

The frame property can now be stated in operator form as

$$AI \le T^*T \le BI, \tag{7}$$

where I is the identity on $\mathcal{H}$. In bra-ket notation,

$$G \equiv T^*T = \int_M d\mu(m)\,|\, h_m \,\rangle\langle\, h_m \,|\,, \tag{8}$$

where the integral is to be interpreted, initially, as converging in the weak operator topology, i.e. as a quadratic form. For a measurable subset N of M, write

$$G(N) \equiv \int_N d\mu(m)\,|\, h_m \,\rangle\langle\, h_m \,|. \tag{9}$$

Proposition 1.2. *If the integral $G(N)$ converges in the strong operator topology of $\mathcal{H}$ whenever N has finite measure, then so does the complete integral representing $G = T^*T$.*

Proof*. Since M is σ–finite, we can choose an increasing sequence $\{M_n\}$ of sets of finite measure such that $M = \cup_n M_n$. Then the corresponding sequence of integrals G_n forms a bounded (by G) increasing sequence of Hermitian operators, hence converges to G in the strong operator topology (see Halmos [1967], problem 94). ∎

If the frame is tight, then $G = AI$ and the above gives a resolution of unity after dividing by A. For non-tight frames, one generally has to do some work to obtain a resolution of unity. The frame condition means that G has a bounded inverse, with

$$B^{-1}I \le G^{-1} \le A^{-1}I. \tag{10}$$

* I thank M. B. Ruskai for suggesting this proof.

Given a function $g(m)$ in $L^2(d\mu)$, we are interested in answering the following two questions: (a) Is $g = Tf$ for some $f \in \mathcal{H}$? (b) If so, then what is f? In other words, we want to:

(a) Find the range $\Re_T \subset L^2(d\mu)$ of the map T.

(b) Find a left inverse S of T, which enables us to reconstruct f from Tf by $f = STf$.

Both questions will be answered if we can explicitly compute G^{-1}. For let

$$P = TG^{-1}T^*\colon L^2(d\mu) \to L^2(d\mu). \tag{11}$$

Then it is easy to see that

$$\begin{aligned} &(a)\ P^* = P \\ &(b)\ P^2 = P \\ &(c)\ PT = T. \end{aligned} \tag{12}$$

It follows that P is the orthogonal projection onto the range of T,

$$P\colon L^2(d\mu) \to \Re_T, \tag{13}$$

for if $g = Tf$ for some f in $\mathcal{H}$, then $Pg = PTf = Tf = g$, and conversely if for some g we have $Pg = g$, then $g = T(G^{-1}T^*g) \equiv Tf$. Thus $\Re_T$ is a closed subspace of $L^2(d\mu)$ and a function $g \in L^2(d\mu)$ is in $\Re_T$ if and only if

$$\begin{aligned} g(m) &= (Pg)(m) = \big(TG^{-1}T^*g\big)(m) \\ &= \langle\, h_m \,|\, G^{-1}T^*g \,\rangle_{\mathcal{H}} \\ &= \langle\, TG^{-1}h_m \,|\, g \,\rangle_{L^2(d\mu)} \\ &= \int_M d\mu(m')\, \overline{\big(TG^{-1}h_m\big)(m')}\, g(m') \\ &= \int_M d\mu(m')\, \overline{\langle\, h_{m'} \,|\, G^{-1}h_m \,\rangle}\, g(m') \\ &= \int_M d\mu(m')\, \langle\, h_m \,|\, G^{-1}h_{m'} \,\rangle\, g(m'). \end{aligned} \tag{14}$$

The function

$$K(m, m') \equiv \langle\, h_m \,|\, G^{-1}h_{m'} \,\rangle \tag{15}$$

therefore has a property similar to the Dirac δ-function with respect to the measure $d\mu$, in that it *reproduces* functions in $\mathfrak{K}_T$. But it differs from the δ-function in some important respects. For one thing, it is bounded by

$$\begin{aligned} |\, K(m,m')\,| \;&=\; |\,\langle\, h_m \,|\, G^{-1} h'_m \,\rangle\,| \\ &\le \|G^{-1}\|\,\|h_m\|\,\|h_{m'}\| \\ &\le A^{-1}\|h_m\|\,\|h_{m'}\| < \infty \end{aligned} \tag{16}$$

for all m and m'. Furthermore, the "test functions" which $K(m,m')$ reproduces form a Hilbert space and $K(m,m')$ defines an integral operator, not merely a distribution, on $\mathfrak{K}_T$. In the applications to relativistic quantum theory to be developed later, M will be a complexification of spacetime and $K(m,m')$ will be holomorphic in m and antiholomorphic in m'.

The Hilbert space $\mathfrak{K}_T$ and the associated function $K(m,m')$ are an example of an important structure called a *reproducing–kernel Hilbert space* (see Meschkowski [1962]), which is reviewed briefly in the next section. $K(m,m')$ is called a *reproducing kernel* for $\mathfrak{K}_T$.

We can thus summarize our answer to the first question by saying that a function $g \in L^2(d\mu)$ belongs to the range of T if and only if it satisfies the *consistency condition*

$$g(m) = \int_M d\mu(m')\, K(m,m') g(m'). \tag{17}$$

Of course, this condition is only useful to the extent that we have information about the kernel $K(m,m')$ or, equivalently, about the operator G^{-1}. The answer to our second question also depends on the knowledge of G^{-1}. For once we know that $g = Tf$ for some $f \in \mathcal{H}$, then

$$f = G^{-1}Gf = G^{-1}T^*\big(Tf\big) = G^{-1}T^*g. \tag{18}$$

Thus the operator

$$S = G^{-1}T^*\colon L^2(d\mu) \to \mathcal{H} \tag{19}$$

is a left inverse of T and we can reconstruct f by

$$\begin{aligned} f = Sg &= G^{-1}T^*Tf \\ &= G^{-1}\int_M d\mu(m)\,|\,h_m\,\rangle\langle\,h_m\,|\,f\,\rangle \\ &= \int_M d\mu(m)\,G^{-1}\,|\,h_m\,\rangle g(m). \end{aligned} \tag{20}$$

This gives f as a linear combination of the vectors

$$h^m \equiv G^{-1}h_m. \tag{21}$$

Note that

$$\int_M d\mu(m)\,|\,h^m\,\rangle\langle\,h^m\,| = G^{-1}GG^{-1} = G^{-1}, \tag{22}$$

therefore the set

$$\mathcal{H}^M \equiv \{h^m\,|\,m \in M\} \tag{23}$$

is also a frame, with frame constants $0 < B^{-1} \leq A^{-1}$. We will call $\mathcal{H}^M$ the frame *reciprocal* to $\mathcal{H}_M$. (In Daubechies [1988a], the corresponding discrete object is called the *dual* frame, but as we shall see below, it is actually a generalization of the concept of reciprocal basis; since the term "dual basis" has an entirely different meaning, we prefer "reciprocal frame" to avoid confusion.)

The above reconstruction formula is equivalent to the resolutions of unity in terms of the pair $\mathcal{H}_M$, $\mathcal{H}^M$ of reciprocal frames:

$$\int_M d\mu(m)\;|\,h^m\,\rangle\langle\,h_m\,| = I = \int_M d\mu(m)\;|\,h_m\,\rangle\langle\,h^m\,|. \tag{24}$$

Corollary 1.3. *Under the assumptions of proposition 1.2, the above resolutions of unity converge in the strong operator topology of* $\mathcal{H}$.

The proof is similar to that of proposition 1.2 and will not be given. The strong convergence of the resolutions of unity is important, since it means that the reconstruction formula is valid *within* $\mathcal{H}$ rather than just weakly. Application to $f = h_k$ for a fixed $k \in M$ gives

$$h_k = \int_M d\mu(m)\,h_m K(m,k), \tag{25}$$

which shows that the frame vectors h_m are in general not linearly independent. The consistency condition can be understood as requiring

the proposed function $g(m)$ to respect the linear dependence of the frame vectors. In the special case when the frame vectors *are* linearly independent, the frames $\mathcal{H}_M$ and $\mathcal{H}^M$ both reduce to bases of $\mathcal{H}$. If $\mathcal{H}$ is separable (which we assume it is), it follows that M must be countable, and without loss in generality we may assume that $d\mu$ is the counting measure on M (re–normalize the h_m's if necessary). Then the above relation becomes

$$h_k = \sum_{m \in M} h_m K(m,k), \tag{26}$$

and linear independence requires that K be the Kronecker δ: $K(m,k) = \delta_m^k$. Thus when the h_m's are linearly independent, $\mathcal{H}_M$ and $\mathcal{H}^M$ reduce to a pair of *reciprocal bases* for $\mathcal{H}$. The resolutions of unity become

$$\sum_{m \in M} |\, h_m \,\rangle\langle\, h^m \,| = I = \sum_{m \in M} |\, h^m \,\rangle\langle\, h_m \,|, \tag{27}$$

and we have the relation

$$h_k = \sum_{m \in M} h^m \langle\, h_m \,|\, h_k \,\rangle \equiv \sum_{m \in M} h^m g_{mk} \tag{28}$$

where

$$g_{mk} \equiv \langle\, h_m \,|\, h_k \,\rangle = \langle\, h^m \,|\, G h_k \,\rangle \tag{29}$$

is an infinite-dimensional version of the *metric tensor*, which mediates between covariant and contravariant vectors. (The operator G plays the role of a *metric operator*.) In this case, $\Re_T = L^2(d\mu) \equiv \ell^2(M)$ and the consistency condition reduces to an identity. The reconstruction formula becomes the usual expression for f as a linear combination of the (reciprocal) basis vectors. If we further specialize to the case of a *tight* frame, then $G = AI$ implies that

$$\langle\, h_m \,|\, h_k \,\rangle = A\, \delta_m^k \qquad \text{and} \qquad \langle\, h^m \,|\, h^k \,\rangle = A^{-1} \delta_k^m, \tag{30}$$

so $\mathcal{H}_M$ and $\mathcal{H}^M$ become *orthogonal* bases. Requiring $A = B = 1$ means that $\mathcal{H}_M = \mathcal{H}^M$ reduce to a single orthonormal basis.

Returning to the general case, we may summarize our findings as follows: $\mathcal{H}_M$, $\mathcal{H}^M$, $K(m,m')$ and $g(m,m') \equiv \langle\, h^m \,|\, G h_{m'} \,\rangle$ are generalizations of the concepts of basis, reciprocal basis, Kronecker

delta and metric tensor to the infinite–dimensional case where, in addition, the requirement of linear independence is dropped. The point is that the all-important reconstruction formula, which allows us to express any vector as a linear combination of the frame vectors, survives under the additional (and obviously necessary) restriction that the consistency condition be obeyed. The useful concepts of orthogonal and orthonormal bases generalize to tight frames and frames with $A = B = 1$, respectively. We will call frames with $A = B = 1$ *normal*. Thus normal frames are nothing but resolutions of unity.

Returning to the general situation, we must still supply a way of computing G^{-1}, on which the entire construction above depends. In some of the examples to follow, G is actually a multiplication operator, so G^{-1} is easy to compute. If no such easy way exist, the following procedure may be used. From $AI \leq G \leq BI$ it follows that

$$-\frac{1}{2}(B-A)I \leq G - \frac{1}{2}(B+A)I \leq \frac{1}{2}(B-A)I. \tag{31}$$

Hence letting

$$\delta = \frac{B-A}{B+A} \qquad \text{and} \qquad c = \frac{2}{B+A}, \tag{32}$$

we have

$$-\delta I \leq I - cG \leq \delta I. \tag{33}$$

Since $0 \leq \delta < 1$ and $c > 0$, we can expand

$$G^{-1} = c\big[I - (I - cG)\big]^{-1} = c\sum_{k=0}^{\infty}(I - cG)^k \tag{34}$$

and the series converges uniformly since

$$\|I - cG\| \leq \delta < 1. \tag{35}$$

The smaller δ, the faster the convergence. For a tight frame, $\delta = 0$ and $cG = I$, so the series collapses to a single term $G^{-1} = c$. Then the consistency condition becomes

$$g(m) = c\int_M d\mu(m')\,\langle\, h_m \,|\, h_{m'}\,\rangle g(m'), \tag{36}$$

and the reconstruction formula simplifies to

$$f = c \int_M d\mu(m)\, h_m\, g(m). \tag{37}$$

If $0 < \delta \ll 1$, the frame is called *snug* (Daubechies [1988a]) and the above formulae merely represent the first terms in the expansions. However, it is found in practice that under certain conditions, this first term gives a good approximation to the reconstruction for snug frames. It appears to be an advantage to have much linear dependence among the frame vectors (precisely that which is impossible when dealing with bases!), so the transformed function $\tilde{f}(m)$ is "oversampled." For such frames, the oversampling seems to compensate for the truncation of the series. A good measure of the amount of linear dependence among the h_m's is the size of the orthogonal complement $\Re_T^\perp$ of the range of T, which is the null space $\mathcal{N}(T^*)$ of T^* since

$$\begin{aligned} g \in \Re_T^\perp &\iff \langle\, g \,|\, Tf \,\rangle = 0 \qquad \forall f \in \mathcal{H} \\ &\iff \langle\, T^*g \,|\, f \,\rangle = 0 \qquad \forall f \in \mathcal{H} \\ &\iff T^*g = 0. \end{aligned} \tag{38}$$

Suppose we are given some function $g(m) \in L^2(d\mu)$ and apply the reconstruction formula to it blindly, without worrying whether the consistency condition is satisfied. That is, consider the vector

$$f \equiv G^{-1}T^*g \tag{39}$$

in $\mathcal{H}$. Since g can be uniquely written as an orthogonal sum $g = g_1 + g_2$ where $g_1 \in \Re_T$ and $g_2 \in \Re_T^\perp = \mathcal{N}(T^*)$, we find that

$$f = G^{-1}T^*g_1 \tag{40}$$

where g_1 does satisfy the consistency condition. This means that applying the reconstruction formula to an *arbitrary* $g \in L^2(d\mu)$ results in a *least–squares approximation* f to a reconstruction, in the sense that the "error" $g_2 \equiv g - g_1 = g - Tf$ has a minimal norm in $L^2(d\mu)$.

For example, suppose we only have information about $\tilde{f}(m)$ for m in some subset Γ of M. Let

$$g(m) = \begin{cases} \tilde{f}(m) & \text{if } m \in \Gamma \\ 0 & \text{if } m \notin \Gamma. \end{cases} \tag{41}$$

Then g does not belong to $\Re_T$, in general, and $f_1 \equiv G^{-1}T^*g$ is the least–squares approximation to the (unknown) vector f in view of our

ignorance. A similar argument applies if M is discrete and f must be reconstructed from Tf by numerical methods. Then we must confine ourselves to a *finite* subset Γ of M. The above procedure then gives a least–squares approximation to f by *truncating* the reconstruction formula to a finite sum over Γ.

A final note: The usual argument in favor of using *bases* rather than overcomplete sets of vectors is that one desires a *unique* representation of functions as linear combinations of basis elements. When a frame is not a basis, i.e. when the frame vectors are linearly dependent, this uniqueness is indeed lost since we may add an arbitrary function $e(m) \in \Re_T^\perp$ to $\tilde{f}(m)$ in the reconstruction formula without changing f ($\Re_T^\perp \neq \{0\}$ since the frame vectors are dependent). However, there is still a unique *admissible* coefficient function, i.e. one satisfying the consistency condition. Moreover, as we shall see, it usually happens in practice that the set M, in addition to being a measure space, has some further structure, and the reproducing kernel $K(m, m')$ preserves this structure. For example, M could be a topological space and K be continuous on $M \times M$, or M could be a differentiable manifold and K be differentiable, or (as will happen in our treatment of relativistic quantum mechanics) M could be a complex manifold and K be holomorphic. Furthermore, K could exhibit a certain boundary– or asymptotic behavior. In such cases, these properties are inherited by all the functions $\tilde{f}(m)$ in $\Re_T$, and then of all possible coefficient functions for a given $f \in \mathcal{H}$, there is only *one* which exhibits the appropriate behavior. In this sense, uniqueness is restored. We will refer to frames with such additional properties as *continuous*, *differentiable*, *holomorphic*, etc.

In addition to properties such as differentiability or holomorphy, the kernels K we will encounter will usually have certain *invariance* properties with respect to some group of transformations acting on M. This, too, will induce a corresponding structure on the function space $\Re_T$.

1.4. Reproducing–Kernel Hilbert Spaces

The function $K(m, m')$ of section 1.3 is an example of a very general structure called a *reproducing kernel*, which we now review briefly since it, too, will play an important role in the chapters to come. The reader interested in learning more about this fascinating subject may consult Aronszajn [1950], Bergman [1950], Meschkowski [1962]

or Hille [1972].

Suppose we begin with an arbitrary set M and a set of functions $g(m)$ on M which forms a Hilbert space $\mathcal{F}$ under some inner product $\langle \cdot | \cdot \rangle$. In section 1.3, M happened to a measure space, $\mathcal{F}$ was $\Re_T$ and the inner product was that of $L^2(d\mu)$. But it is important to keep in mind that the exact form of the inner product in $\mathcal{F}$ does not need to be specified, as far as the general theory of reproducing kernels is concerned. Suppose we are given a complex–valued function $K(m, m')$ on $M \times M$ with the following two properties:

1. For every $m \in M$, the function $K_m(m') \equiv K(m', m)$ belongs to $\mathcal{F}$.
2. For every $m \in M$ and every $g \in \mathcal{F}$, we have

$$g(m) = \langle K_m \,|\, g \rangle. \tag{1}$$

Then $\mathcal{F}$ is called a *reproducing–kernel Hilbert space* and $K(m, m')$ is called its reproducing kernel. Some properties of K follow immediately from the definition. For example, since $K_m \in \mathcal{F}$, property (2) implies that

$$K(m', m) = K_m(m') = \langle K_{m'} \,|\, K_m \rangle. \tag{2}$$

Thus K must satisfy

$$\begin{aligned} &(a) \quad \overline{K(m, m')} = K(m', m) \\ &(b) \quad K(m, m) = \|K_m\|^2 \geq 0 \quad \forall m \in M \\ &(c) \quad |K(m, m')| \leq \|K_m\| \, \|K_{m'}\|. \end{aligned} \tag{3}$$

One of the most important and useful aspects of reproducing–kernel Hilbert spaces is the fact that the kernel function itself virtually generates the whole structure. For example, the function $L(m) \equiv \|K_m\|$ dominates every $g(m)$ in $\mathcal{F}$ in the sense that

$$|g(m)| = |\langle K_m \,|\, g \rangle| \leq \|K_m\| \, \|g\| = \|g\| L(m) \tag{4}$$

by the Schwarz inequality. In particular, all the functions in $\mathcal{F}$ inherit the boundednes and growth properties of K. If K has singularities, then *some* functions in $\mathcal{F}$ have similar singularities.

From the above it follows that for fixed $m_0 \in M$,

$$\sup_{\|g\|=1} |g(m_0)| = L(m_0), \tag{5}$$

the supremum being attained by $g_{m_0}(m) = K_{m_0}(m)/L(m_0)$, and this is, up to an overall phase factor, the only function which attains the supremum. This fact has an important interpretation when M is the classical phase space of some system, $\mathcal{F}$ represents the Hilbert space of the corresponding quantum mechanical system and $|g(m)|^2$ is the probability density in the quantum mechanical state g of finding the system in the classical state m. The above inequality then shows that the state which maximizes the probability of being at m_0 is uniquely determined (up to a phase factor) as g_{m_0}. In other words, g_{m_0} is a wave packet which is in some sense optimally localized at m_0 in phase space.

Another example of how the kernel function K embodies the properties of the entire Hilbert space $\mathcal{F}$ is the following: If the basic set M has some additional structure, such as being a measure space (as it was in the setting of frame theory) or a topological space or a C^k, C^∞, real–analytic or complex manifold, and if $K(m, m')$, as a function of m, preserves this structure (that is, $K(m, m')$ is measurable, continuous, C^k, C^∞, real–analytic or holomorphic in m, respectively), then every function $g(m)$ in $\mathcal{F}$ has the same property.

The question arises: If we are given a Hilbert space $\mathcal{F}$ whose elements are all functions on M, how do we know whether this space posesses a reproducing kernel? Clearly a *necessary* condition for K to exist is that

$$|\, f(m)\,| \;=\; |\,\langle\, K_m \,|\, f\,\rangle\,| \;\leq\; \|K_m\|\,\|f\| \tag{6}$$

for all $f \in \mathcal{F}$ and all $m \in M$. But this means that for every *fixed* m, the map $E_m\colon \mathcal{F} \to \mathbb{C}$ defined by

$$E_m(f) = f(m) \tag{7}$$

is a bounded linear functional on $\mathcal{F}$. E_m is called the *evaluation map* at m. By the Riesz representation theorem, every bounded linear functional E on $\mathcal{F}$ must have the form $E(f) = \langle\, e \,|\, f\,\rangle$ for a unique vector $e \in \mathcal{F}$. Hence there exists a unique $e_m \in \mathcal{F}$ such that

$$f(m) = \langle\, e_m \,|\, f\,\rangle \tag{8}$$

for all $f \in \mathcal{F}$. Since also $f(m) = \langle\, K_m \,|\, f\,\rangle$, we conclude that $e_m = K_m$. Thus it follows that given *only* $\mathcal{F}$ such that all the evaluation maps E_m are bounded (not necessarily uniformly in m), we can *construct* a reproducing kernel by

$$K(m, m') = \langle\, e_m \,|\, e_{m'} \,\rangle. \tag{9}$$

That is, $\mathcal{F}$ *posesses a reproducing kernel if and only if all the evaluation maps are bounded.*

In all of the applications we will encounter, the set M will have a structure beyond those mentioned so far: it will be a *Lie group* G or a *homogeneous space* of G. That is, each $g \in G$ *acts* on M as an invertible transformation preserving whatever other structure M may have such as continuity, differentiability, etc, and these tranformations form the group G under composition. Let us denote the action of g on M by $m \mapsto mg$ (i.e., G acts from the *right*). Then the operator

$$\begin{aligned} T_g &: \mathcal{F} \to \mathcal{F} \\ (T_g f)(m) &= f(mg) \end{aligned} \tag{10}$$

gives a *representation* of G on $\mathcal{F}$, since $T_g T_{g'} = T_{gg'}$. From $f(m) = \langle\, K_m \,|\, f \,\rangle$ it now follows that

$$K_{mg} = T_g^*\, K_m, \tag{11}$$

hence

$$K(m'g, mg) = \langle\, K_{m'g} \,|\, K_{mg} \,\rangle = \langle\, K_{m'} \,|\, T_g\, T_g^*\, K_m \,\rangle. \tag{12}$$

Therefore *the reproducing kernel* $K(m', m)$ *is invariant under the action of* G *if and only if all the operators* T_g *are unitary, i.e. the representation* $g \mapsto T_g$ *is unitary.*

The group G usually appears in applications as a natural set of operations such as translations in space and time (evolution), changes of reference frame or coordinate system, dilations and frequency shifts (especially useful in signal analysis), etc. Invariance under G then means that the transformed objects (wave functions, signals) form a description of the system equivalent to the original one, i.e. that the transformation changes nothing of physical significance.

Finally, let us note that the concept of generalized frame, as defined in the previous section, is a special case of a reproducing–kernel Hilbert space. For given such a frame $\{h_m\}$, the function $K(m', m) = \langle\, h_{m'} \,|\, G^{-1} h_m \,\rangle$ is a reproducing kernel for the space $\Re_T$ since (a):

$$K_m(m') \equiv K(m', m) = \big(T(G^{-1} h_m)\,\big)\,(m') \tag{13}$$

shows that $K_m = T(G^{-1}h_m)$ belongs to $\Re_T$, and (b):

$$\tilde{f}(m) = \int_M d\mu(m')\, K(m,m')\tilde{f}(m') = \langle\, K_m \,|\, \tilde{f}\,\rangle_{\Re_T}. \qquad (14)$$

1.5. Windowed Fourier Transforms

The coherent states of section 2 form a (holomorphic) frame. I now want to give some other examples of frames, in order to develop a better feeling for this concept. The frames to be constructed in this section turn out to be closely related to the coherent states, but have a distinct "signal processing" flavor which will lend some further depth to our understanding of phase-space localization.

For simplicity, we begin with the study of functions $f(t)$ of a single real variable which, for motivational purposes, will be thought of as "time." Although in the model to be built below f is real-valued, we allow it to be complex-valued since our eventual applications will be quantum–mechanical. The same goes for the window function h. The extension to functions of several variables, such as wave functions or physical fields in spacetime, is straightforward and will be indicated later. We think of $f(t)$ as a "time–signal," such as the voltage going into a speaker or the pressure on an eardrum. We are interested in the frequency content of this signal. The standard thing to do is to find its Fourier coefficients (if f is periodic) or its Fourier transform (if it is not periodic). But if f represents, say, a symphony, this approach is completely inappropriate. We are not interested in the *total* amplitude

$$\hat{f}(\nu) = \int_{-\infty}^{\infty} dt\, e^{2\pi i\nu t} f(t) \qquad (1)$$

in f of each frequency ν. For one thing, both we and the musicians would be long gone before we got to enjoy the music. Moreover, the musical content of the signal, though coded into $\hat{f}$, would be as inaccessible to us as it is in $f(t)$ itself. Rather, we want to analyze the frequency content of f in *real* time. At each instant $t = s$ we hear a "spectrum" $\tilde{f}(\nu, s)$. To accomplish this, let us make a simple (though not very realistic) linear model of our auditory system. We will speak of the "ear," though actually the frequency analysis appears to be performed partly by the nervous system as well (Roederer [1975]).

The ear's output at time s depends on the input $f(t)$ for t in some interval $s - \tau \le t \le s$, where τ is a lag–time characteristic of the ear. In analyzing $f(t)$ in this interval, the ear may give different weights to different parts of $f(t)$. Thus the signal to be analyzed for output at time s may be modeled as

$$f_s(t) = \overline{h(t-s)}\, f(t) \tag{2}$$

where $\overline{h(t-s)}$ is the weight assigned to $f(t)$. (As noted above, we are allowing h to be complex–valued for future applications, though here it should be real–valued; the bar means complex conjugation.) The function $h(t)$ is characteristic of the ear, with support in the interval $-\tau \le t \le 0$. Such functions are known in communication theory as *windows*.

Having "localized" the signal around time s, we now analyze its frequency content by taking the Fourier transform:

$$\tilde{f}(\nu, s) = \hat{f}_s(\nu) = \int_{-\infty}^{\infty} dt\, e^{2\pi i \nu t}\, \overline{h(t-s)}\, f(t). \tag{3}$$

This function, representing our "dynamical spectrum," is called a *windowed Fourier transform* of f (Daubechies [1988a]). If the window is flat ($h(t) \equiv 1$), $\tilde{f}(\nu, s)$ reduces to the ordinary Fourier transform. On the other hand, if $f(t)$ is a "unit impulse" at $t = 0$, i.e. $f(t) = \delta(t)$, then $\tilde{f}(\nu, s) = \overline{h(-s)}$. Hence $\overline{h(-s)}$ is the "impulse response" (Papoulis [1962]) of the ear. Note that

$$\tilde{f}(0, s) = \int_{-\infty}^{\infty} dt\, \overline{h(t-s)}\, f(t) \tag{4}$$

is nothing but the convolution of f with the impulse response. Thus the windowed Fourier transform is a marriage between the Fourier transform and convolution. Letting

$$h_{\nu,s}(t) = e^{-2\pi i \nu t}\, h(t-s), \tag{5}$$

we have

$$\tilde{f}(\nu, s) = \langle\, h_{\nu,s} \,|\, f \,\rangle, \tag{6}$$

the inner product being in $L^2(\mathbb{R})$. As our notation implies, we want to make a frame indexed by the set

$$M = \{(\nu, s) \,|\, \nu, s \in \mathbb{R}\} \approx \mathbb{R}^2, \tag{7}$$

that is, the *time-frequency plane*. M corresponds to the "phase space" in quantum mechanics in the sense that if t were the position coordinate, then ν would be the wavenumber and $2\pi\hbar\nu$ would be the momentum. Since we expect all times and all frequencies to be equally important, let us guess that an appropriate measure on M is the Lebesgue measure, $d\mu(\nu, s) = ds\, d\nu$. Now

$$\tilde{f}(\nu, s) = \big(\bar{h}_s f\big)\hat{}\,(\nu) \tag{8}$$

where $\bar{h}_s(t) = \overline{h(t-s)}$ is the translated window and $\hat{}$ denotes the Fourier transform. Thus for "nice" signals (say, in the space of Schwartz test functions), Plancherel's theorem gives

$$\begin{aligned} \int_{\mathbb{R}^2} ds\, d\nu \;|\, \tilde{f}(\nu, s)\,|^{\,2} &= \int ds \int d\nu \;|\, (\bar{h}_s f)\hat{}\,(\nu)\,|^{\,2} \\ &= \int ds \int dt \;|\, \overline{h_s(t)} f(t)\,|^{\,2} \\ &= \int dt \;|\, f(t)\,|^{\,2} \int ds \;|\, h(t-s)\,|^{\,2} \\ &= \|h\|^2\, \|f\|^2, \end{aligned} \tag{9}$$

where both norms are those of $L^2(\mathbb{R})$. This shows that the family of vectors $h_{\nu,s}$ is indeed a tight frame, with frame constants $A = B = \|h\|^2$. Now $h_{\nu,s} = \exp(-2\pi i\nu t) h_s(t)$ is just a translation of h_s in *frequency*, that is,

$$\hat{h}_{\nu,s}(\nu') = \hat{h}_s(\nu' - \nu). \tag{10}$$

Thus if the Fourier transform $\hat{h}(\nu)$ of the basic window is concentrated near $\nu = 0$ (which, among other things, means that $h(t)$ must be fairly smooth, since any discontinuities or sharp edges introduce high-frequency components), then that of $h_{\nu,s}$ is concentrated near $\nu' = \nu$. In other words, $h_{\nu,s}$ is actually a window in *phase space* or time–frequency. The fact that these windows form a tight frame means that signals may be equally well represented in time–frequency as in time. The resolution of unity is

$$\|h\|^{-2} \int_{\mathbb{R}^2} ds\, d\nu \; | \, h_{\nu,s} \, \rangle \langle \, h_{\nu,s} \, | \; = I, \tag{11}$$

and if the consistency condition is satisfied the reconstruction formula gives

$$f(t) = \|h\|^{-2} \int_{\mathbb{R}^2} ds\, d\nu \, e^{-2\pi i \nu t} \, h(t-s) \, \tilde{f}(\nu, s). \tag{12}$$

As in section 1.2, we denote by $\mathcal{H}_M$ the frame formed by the $h_{\nu,s}$'s.

Incidentally, the windowed Fourier transform is quite symmetrical with respect to the interchange of time and frequency. It can be re–written as

$$\begin{aligned} \tilde{f}(\nu, s) &= \left(\bar{h}_s \, f\right)\hat{}\,(\nu) \\ &= \left(\hat{\bar{h}}_s * \hat{f}\right)(\nu) \\ &= e^{-2\pi i \nu s} \int d\nu' \, e^{2\pi i \nu' s} \, \overline{\hat{h}(\nu' - \nu)} \, \hat{f}(\nu'), \end{aligned} \tag{13}$$

where $\hat{h}$ now plays the role of a window in frequency used to localize $\hat{f}$. As will be seen in the next chapter, the reason for this symmetry is that windowed Fourier transforms, like the canonical coherent states, are closely related to the Weyl–Heisenberg group, which treats time and frequency in a symmetrical fashion. (This is rooted in symplectic geometry.)

The $h_{\nu,s}$'s are highly redundant. We want to find discrete subsets of them which still form a frame, that is we want *discrete subframes*. The following construction is taken from Kaiser [1978c, 1984a]. Let $T > 0$ be a fixed time interval and suppose we "sample" the output signal $\tilde{f}(\nu, s)$ only at times $s = nT$ where n is an integer. To discretize the frequency as well, note that the localized signal $f_{nT}(t) = \bar{h}_{nT}(t) f(t)$ has compact support in the interval $nT \; - \tau \leq \; t \leq \; nT$, hence we can expand it in a Fourier series

$$f_{nT}(t) = \sum_m e^{-2\pi i m F t} \, c_{mn} \tag{14}$$

where $F = 1/\tau$ and

$$\begin{aligned} c_{mn} &= F \int_{nT-\tau}^{nT} dt\, e^{2\pi i m F t}\, f_{nT}(t) \\ &= F \int_{nT-\tau}^{nT} dt\, e^{2\pi i m F t}\, \overline{h(t-nT)}\, f(t) \\ &= F\, \tilde{f}(mF, nT). \end{aligned} \tag{15}$$

The only problem with this representation of f_{nT} is that it only holds in the interval $nT - \tau \leq t \leq nT$, since f_{nT} vanishes outside this interval while the Fourier series is periodic. To force equality at all times, we multiply both sides by $\tau h_{nT}(t)$:

$$\begin{aligned} \tau h_{nT}(t)\, f_{nT}(t) &= \tau\, |h_{nT}(t)|^2\, f(t) \\ &= \sum_m e^{-2\pi i m F t}\, h_{nT}(t)\, \tilde{f}(mF, nT) \\ &= \sum_m h_{mF,nT}(t)\, \langle\, h_{,mF,nT} \,|\, f \,\rangle. \end{aligned} \tag{16}$$

Attempting to recover the entire signal rather than just pieces of it, we now sum both sides with respect to n:

$$\tau \Big(\sum_n |h(t-nT)|^2 \Big) f(t) = \sum_{n.m} h_{mF,nT}(t) \langle\, h_{nT,mF} \,|\, f \,\rangle. \tag{17}$$

Recovery of $f(t)$ is possible provided that the sum on the left converges to a function

$$g(t) \equiv \tau \sum_n |h(t-nT)|^2 \tag{18}$$

which is bounded above and below by positive constants:

$$0 < A \leq g(t) \leq B. \tag{19}$$

In that case, we have

$$\sum_{n,m} |\, \langle\, h_{mF,nT} \,|\, f \,\rangle\, |^{\,2} = \int dt\, g(t) |f(t)|^2 \tag{20}$$

and hence the subset

$$\mathcal{H}_M^{T,F} \equiv \{h_{mF,nT} \,|\, m, n \in \mathbb{Z}\} \tag{21}$$

forms a discrete subframe of $\mathcal{H}_M$ with frame constants A and B. This subframe is in general not tight, and the operator G is just multiplication by $g(t)$, so finding G^{-1} presents no problem in this case. The reconstruction formula is

$$f(t) = g(t)^{-1} \sum_{n,m \in \mathbb{Z}} h_{mF,nT}(t) \tilde{f}(mF, nT). \tag{22}$$

We are therefore able to recover the signal by "sampling" it in phase space at time intervals $\Delta t = T$ and frequency intervals $\Delta\nu = F$. What about the uncertainty principle? It is hiding in the condition that the discrete subset h_{mFnT} forms a frame! For we cannot satisfy $g(t) \geq A > 0$ unless the supports of $h(t)$ and $h(t-T)$ overlap, which implies that $T \leq \tau$, or $\Delta t \Delta\nu = T/\tau \leq 1$. In quantum mechanics, the radian frequency is related to the energy by $E = 2\pi\hbar\nu$, so the above condition is

$$\Delta t \, \Delta E \leq 2\pi\hbar. \tag{23}$$

This looks like the uncertainty principle going "the wrong way." The intuitive explanation for this has been discussed at the end of section 1.3.

Notice that the closer we choose T to τ, the more difficult it is for the window function to be smooth. In the limiting case $T = \tau$, $h(t)$ must be discontinuous if the above frame condition is to be obeyed. As noted earlier, this means that its Fourier transform $\hat{h}(\nu)$ can no longer be concentrated near $\nu = 0$, so the frequency resolution of the samples suffers. In concrete terms, this means that whereas for "nice" windows $h(t)$ we may hope to get a good approximation to the reconstruction formula by truncating the double sum after an appropriate finite number of terms, this can no longer be expected when $T \to \tau$. In other words, it pays to oversample! "Appropriate" here means that we cover most of the area in the time–frequency plane where $\tilde{f}$ lives. Clearly, if the sampling is done only for $|n| \leq N$, we cannot expect to recover $f(t)$ outside the interval $-NT - \tau \leq t \leq NT$. If $\hat{h}(\nu)$ is nicely peaked around $\nu = 0$, say with a spread of $\Delta\nu$, and the signal $f(t)$ is (approximately) "band–limited," so that $\hat{f}(\nu) \approx 0$ for $|\nu| \geq W$, then we can expect to get a good approximation to $f(t)$ by truncating the sum with $m \leq M$, where

M is chosen to satisfy $MF > W + \Delta\nu$. On the other hand, we may actually not be interested in recovering the exact original signal $f(t)$. If we are only interested in a particular time interval and a particular frequency interval (say, to eliminate some high–frequency noise), then the appropriate truncation would include only an area in the time–frequency plane which is slightly larger than our area of interest. If we sample a given signal f only in a finite subset Γ of our lattice (as we are bound to do in practice) and then apply the reconstruction formula, truncating the sum by restricting it to Γ, then the result is a least–squares approximation f_1 to the original signal.

On a philosophical note, suppose we are given an arbitrary signal without any idea of the type of information it carries. The choice of a window $h(t)$ then defines a *scale* in time, given by τ, and it is this scale that distinguishes what will be perceived as frequency and what as time. Thus, variations of $f(t)$ in time intervals much smaller than τ are reflected in the ν–behavior of $\tilde{f}$, while variations at scales much larger than τ survive in the s–dependence of $\tilde{f}$. What an elephant perceives as a tone may appear as a rhythm to a mouse.

Finally, I wish to compare the above reconstruction with the Nyquist/Shannon sampling theorem (Papoulis [1962]), which gives a reconstruction for band-limited signals ($\hat{f}(\nu) \approx 0$ for $|\nu| \geq W$) in terms of samples of $f(t)$ taken at times $t = n/2W$ for integer n. Note, first of all, that in the time–frequency reconstruction formula above, it was not necessary to assume that $f(t)$ is band-limited. In fact, the formula applies to all square–integrable signals. But suppose that $f(t)$ *is* band–limited as above, and choose a window $h(t)$ such that $\hat{h}(\nu)$ is concentrated around an interval of width $\Delta\nu$ about the origin, with $\Delta\nu < W$. Then we expect $\tilde{f}(\nu, s) \approx 0$ for $|\nu| \geq W + \Delta\nu$. In order to reduce the double sum in our reconstruction formula to a single sum over n as in the Nyquist theorem, choose $F = W + \Delta\nu$. Then $\tilde{f}(mF, s) \approx 0$ whenever $m \neq 0$. To apply the reconstruction formula, we must still choose a time interval $T < \tau$. By the uncertainty principle, $\tau\Delta\nu > 1/2$. The above condition on $\Delta\nu$ therefore implies that $\tau > 1/2W$. It would thus seem that we could get away with a slightly larger sampling interval T than the Nyquist interval $T_N = 1/2W$. Our reconstruction formula reduces to

$$f(t) \approx g(t)^{-1} \sum_n h(t - nT)\, \tilde{f}(0, nT). \tag{24}$$

The smaller the ratio $\Delta\nu/W$, the better this approximation is likely

to be. But a small $\Delta\nu$ means a large τ, hence the samples $\tilde{f}(0, nT)$ are smeared over a large time interval.

1.6. Wavelet Transforms

The frame vectors for windowed Fourier transforms were the wave packets

$$h_{\nu,s}(t) = e^{-2\pi i\nu t}h(t-s). \tag{1}$$

The basic window function $h(t)$ was assumed to vanish outside of the interval $-\tau < t < 0$ and to be reasonably smooth with no steep slopes, so that its Fourier transform $\hat{h}(\nu)$ was also centered in a small interval about the origin. Of course, since $h(t)$ has compact support, $\hat{h}(\nu)$ is the restriction to $\mathbb{R}$ of an entire function and hence cannot vanish on any interval, much less be of compact support. The above statement simply means that $\hat{h}(\nu)$ decays rapidly outside of a small interval containing the origin. At any rate, the factor $\exp(-2\pi i\nu t)$ amounted to a translation of the window in frequency, so that $h_{\nu,s}$ was a "window" in the time–frequency plane centered about (ν, s). Hence the frequency components of $f(t)$ were picked out by means of rigidly translating the basic window in both time and frequency. (It is for this reason that the windowed Fourier transform is associated to the Weyl–Heisenberg group, which is exactly the group of all translations in phase space amended with the multiplication by phase factors necessary to close the Lie algebra, as explained in chapter 3.) Consequently, $h_{\nu,s}$ has the same width τ for all frequencies, and the number of wavelengths admitted for analysis is $\nu\tau$. For low frequencies with $\nu\tau \ll 1$ this is inadequate since we cannot gain any meaningful frequency information by looking at a small fraction of a wavelength. For high frequencies with $\nu\tau \gg 1$, too many wavelengths are admitted. For such waves, a time-interval of duration τ seems infinite, thus negating the sense of "locality" which the windowed Fourier transform was designed to achieve in the first place. This deficiency is remedied by the *wavelet transform*. The window $h(t)$, called the *basic wavelet*, is now *scaled* to accomodate waves of different frequencies. That is, for $a \neq 0$ let

$$h_{a,s}(t) = |a|^{-1/2}h\left(\frac{t-s}{a}\right). \tag{2}$$

The factor $|a|^{-1/2}$ is included so that

$$\|h_{a,s}\|^2 \equiv \int_{-\infty}^{\infty} dt \mid h_{a,s}(t) \mid^2 = \|h\|^2. \tag{3}$$

The necessity of using negative as well as positive values of a will become clear as we go along. It will also turn out that h will need to satisfy a technical condition. Again, we think of both h and f as real but allow them to be complex. The wavelet transform is now defined by

$$\tilde{f}(a,s) = \langle\, H_\alpha^* \mid f \,\rangle = \int_{-\infty}^{\infty} dt\, |a|^{-1/2} \overline{h\left(\frac{t-s}{a}\right)} f(t). \tag{4}$$

Before proceeding any further, let us see how the wavelet transform localizes signals in the time–frequency plane. The localization in time is clear: If we assume that $h(t)$ is concentrated near $t = 0$ (though it will no longer be convenient to assume that h has compact support), then $\tilde{f}(a,s)$ is a weighted average of $f(t)$ around $t = s$ (though the weight function need not be positive, and in general may even be complex). To analyze the frequency localization, we again want to express $\tilde{f}$ in terms of the Fourier transforms of h and f. This is possible because, like the windowed Fourier transform, the wavelet transform involves rigid time–translations of the window, resulting in a convolution–like expression. The "impulse response" is now (setting $f(t) = \delta(t)$)

$$g_a(s) = |a|^{-1/2} \overline{h(-s/a)}, \tag{5}$$

and we have

$$\tilde{f}(a,s) = \big(g_a * f\big)(s) = \big(\hat{g}_a \hat{f}\big)\check{}\,(s), \tag{6}$$

with

$$\hat{g}_a(\nu) = \int_{-\infty}^{\infty} dt\, e^{2\pi i \nu t} |a|^{-1/2} \overline{h(-t/a)} = |a|^{1/2} \overline{\hat{h}(a\nu)}. \tag{7}$$

Later we will see that discrete tight frames can be obtained with certain choices of $h(t)$ whose Fourier transforms have compact support in a frequency interval interval $\alpha \le \nu \le \beta$. Such functions (or, rather, the operations of convolutiong with them) are called *bandpass filters* in communication theory, since the only frequency components in $f(t)$ to survive are those in the "band" $[\alpha, \beta]$. Then the above expression

shows that $\tilde{f}(a,s)$ depends only on the frequency component of $f(t)$ in the band $\alpha/a \le \nu \le \beta/a$ (if $a > 0$) or $\beta/a \le \nu \le \alpha/a$ (if $a < 0$). Thus frequency localization is achieved by *dilations* rather than translations in frequency space, in contrast to the windowed Fourier transform. At least from the point of view of audio signals, this actually seems preferable since it appears to be frequency ratios, rather than frequency differences, which carry meaning. For example, going up an octave is achieved by doubling the frequency. (However, frequency differences do play a role in connection with beats and also in certain non–linear phenomena such as difference tones; see Roederer [1975].) Let us now try to make a continuous frame out of the vectors $h_{a,s}$. This time the index set is

$$M = \{(a,s) \,|\, a \neq 0, s \in \mathbb{R}\} \approx \mathbb{R}^* \times \mathbb{R} \tag{8}$$

where $\mathbb{R}^*$ denotes the group of non-zero real numbers under multiplication. M is the *affine group* of translations and dilations of the real line, $t' = at + s$, and this fact will be recognized as being very important in chapter 3. But for the present we use a more pedestrian approach to obtain the central results. This will make the power and elegance of the group–theoretic approach to be introduced later stand out and be appreciated all the more. At this point we only make the safe assumption that the measure $d\mu$ on M is invariant under time translations, i.e. that

$$d\mu(a,s) = \rho(a) da\, ds \tag{9}$$

where $\rho(a)$ is an as yet undetermined density on $\mathbb{R}^*$. Then, using Plancherel's theorem,

$$\begin{aligned}
\int_M d\mu(a,s) \,|\, \tilde{f}(a,s) \,|^2 &= \int_M \rho(a)\, da\, ds \,|\, (\hat{g}_a \hat{f})\check{}(s) \,|^2 \\
&= \int_{\mathbb{R}^*} \rho(a) da \int_{\mathbb{R}} d\nu \,|\, \hat{g}_a(\nu) \,|^2 \,|\, \hat{f}(\nu) \,|^2 \\
&= \int_{\mathbb{R}^*} \rho(a) da \int_{\mathbb{R}} d\nu \,|a|\, |\, \hat{h}(a\nu) \,|^2 \,|\, \hat{f}(\nu) \,|^2 \\
&= \int_{-\infty}^{\infty} d\nu\, H(\nu) \,|\, \hat{f}(\nu) \,|^2 \\
&= \langle\, \hat{f} \,|\, H \hat{f} \,\rangle_{L^2(\mathbb{R})},
\end{aligned} \tag{10}$$

where

$$\begin{aligned} H(\nu) &= \int_{\mathbb{R}^*} \rho(a) \cdot |\, a \,| \, da \, |\, \hat{h}(a\nu) \,|^{\,2} \\ &= \int_0^\infty a da \, \left[\rho(a) \, |\, \hat{h}(a\nu) \,|^{\,2} + \rho(-a) \, |\, \hat{h}(-a\nu) \,|^{\,2} \right]. \end{aligned} \tag{11}$$

The frame condition is therefore $A \leq H(\nu) \leq B$ for some positive constants A and B. To see its implications, we analyze the cases $\nu > 0$ and $\nu < 0$ separately. If $\nu > 0$, let $\xi = a\nu$. Then

$$H(\nu) = \nu^{-2} \int_0^\infty \xi d\xi \, \left[\rho(\xi/\nu) \, |\, \hat{h}(\xi) \,|^{\,2} + \rho(-\xi/\nu) \, |\, \hat{h}(-\xi) \,|^{\,2} \right]. \tag{12}$$

If $\nu < 0$, let $\xi = -a\nu$. Then

$$H(\nu) = \nu^{-2} \int_0^\infty \xi d\xi \, \left[\rho(-\xi/\nu) \, |\, \hat{h}(-\xi) \,|^{\,2} + \rho(\xi/\nu) \, |\, \hat{h}(\xi) \,|^{\,2} \right], \tag{13}$$

giving the same expression. Therefore the frame condition requires that

$$\rho(\xi/\nu) = O\big((\xi/\nu)^{-2}\big) \quad \text{as} \quad \xi/\nu \to \pm\infty, \tag{14}$$

unless $\hat{h}(\nu)$ vanishes in a neighborhood of the origin. Note that the above expression for $H(\nu)$ shows that if both $\rho(a)$ and $\hat{h}(\nu)$ vanish for negative arguments then $H(\nu) \equiv 0$ and no frame exists. Hence to support general complex-valued windows (such as bandpass filters for a positive–frequency band), it is necessary to include negative as well as positive scale factors a.

The general case, therefore, is that we get a (generalized) frame whenever $\rho(a)$ and $h(t)$ are chosen such that $0 < A \leq H(\nu) \leq B$ is satisfied. The "metric operator" G and its inverse are given in terms of Fourier transforms by

$$Gf = \big(H\hat{f}\big)\check{} \qquad \text{and} \qquad G^{-1}f = \big(H^{-1}\hat{f}\big)\check{}. \tag{15}$$

Since G is no longer a multiplication operator in the time domain (as it was in the case of the discrete frame we constructed from the

windowed Fourier transforms), the action of G^{-1} is more complicated. It is preferable, therefore, to specialize to tight frames. This requires that $H(\nu)$ be constant, so the asymptotic conditions on ρ reduce to the requirement that ρ be piecewise continuous:

$$\rho(a) = \begin{cases} c^+/a^2, & \text{if } a > 0 \\ c^-/a^2, & \text{if } a < 0 \end{cases} \tag{16}$$

where c^+ and c^- are non-negative constants (not both zero). Then for $\nu > 0$,

$$H(\nu) = \int_0^\infty \frac{d\xi}{\xi} \left[c^+ \,|\, \hat{h}(\xi) \,|^{\,2} + c^- \,|\, \hat{h}(-\xi) \,|^{\,2} \right] \tag{17}$$

and for $\nu < 0$,

$$H(\nu) = \int_0^\infty \frac{d\xi}{\xi} \left[c^- \,|\, \hat{h}(\xi) \,|^{\,2} + c^+ \,|\, \hat{h}(-\xi) \,|^{\,2} \right]. \tag{18}$$

Thus $H(\nu) = A = B$ requires either that $|\,\hat{h}(\xi)\,|^{\,2} = |\,\hat{h}(-\xi)\,|^{\,2}$ (which holds if $h(t)$ is real) or that $c^+ = c^-$. Since we want to accomodate complex wavelets, we assume the latter condition. Then we have

$$A = B = c^+ \int_{\mathbb{R}^*} \frac{d\xi}{|\xi|} \left|\hat{h}(\xi)\right|^2. \tag{19}$$

We have therefore arrived at the measure

$$d\mu(a, s) = c^+ da\, ds/a^2 \tag{20}$$

for tight frames, which coincides with the measure suggested by group theory (see chapter 3). In addition, we have found that the basic wavelet h must have the property that

$$c_h \equiv \int_{\mathbb{R}^*} \frac{d\xi}{|\xi|} \left|\hat{h}(\xi)\right|^2 < \infty. \tag{21}$$

In that case, $h(t)$ is said to be *admissible*. This condition is also a special case of a group–theoretic result, namely that we are dealing with a *square–integrable representation* of the appropriate group (in this case, the affine group $\mathbb{R}^* \times \mathbb{R}$). To summarize, we have constructed a continuous tight frame of wavelets $h_{a,s}$ provided the basic wavelet is admissible. The corresponding resolution of unity is

$$c_h^{-1} \int_{\mathbb{R}^* \times \mathbb{R}} \frac{da\, ds}{a^2} \,|\, h_{a,s} \,\rangle\langle\, h_{a,s} \,|\, = I. \tag{22}$$

The associated reproducing–kernel Hilbert space $\mathfrak{K}_K$ is the space of functions $(Kf)(a,s) = \langle\, h_{a,s} \,|\, f \,\rangle \equiv \tilde{f}(a,s)$ depending on the scale parameter a as well as the time coordinate s. As $a \to 0$, $h_{a,s}$ becomes peaked around $t = s$ and

$$\tilde{f} \sim |a|^{1/2} c f(s) \tag{23}$$

where $c = \int \overline{h(u)}\, du$. The transformed signal $\tilde{f}$ is a smoothed–out version of f and a serves as a *resolution parameter*.

Ultimately, all computations involve a finite number of operations, hence as a first step it would be helpful to construct a *discrete* subframe of our continuous frame. Toward this end, choose a fundamental scale parameter $a > 1$ and a fundamental time shift $b > 0$. We will consider the discrete subset of dilations and translations

$$D = \{(a^m, na^m b) \,|\, m, n \in \mathbb{Z}\} \subset \mathbb{R}^* \times \mathbb{R}. \tag{24}$$

Note that since $a^m > 0$ for all m, only positive dilations are included in D, contrary to the lesson we have learned above. This will be remedied later by considering $\overline{h(t)}$ along with $h(t)$. Also, D is not a subgroup of $\mathbb{R}^* \times \mathbb{R}$, as can be easily checked. The wavelets parametrized by D are

$$h_{mn} = a^{-m/2}\, h\left(\frac{t - na^m b}{a^m}\right) = a^{-m/2} h(a^{-m}t - nb). \tag{25}$$

To see that this is exactly what is desired, suppose $\hat{h}(\nu)$ is concentrated on an interval around $\nu = F$ (i.e., $\hat{h}$ is a band–pass filter). Then $\hat{h}_{mn}$ is concentrated around $\nu = F/a^m$. For given integer m, the "samples"

$$f_{mn} \equiv \langle\, h_{mn} \,|\, f \,\rangle, \quad n \in \mathbb{Z} \tag{26}$$

therefore represent (in discrete "time" n) the behavior of that part of the signal $f(t)$ with frequencies near F/a^m. If $m \gg 1$, f_{mn} will vary slowly with n, and if $m \ll -1$, it will vary rapidly with n (if $f(t)$ has frequency components with $\nu \sim F/a^m$). Now the time–samples are separated by the interval $\Delta_m t = a^m b$, so the *sampling rate*

$$R_m = 1/a^m b \tag{27}$$

is automatically adjusted to the frequency range $\nu \sim F/a^m$ of the output signal $\{f_{mn} \,|\, n \in \mathbb{Z}\}$: The high–frequency components get sampled proportionately more often. This is an example of a topic in mathematics which has recently attracted intense activity under the banner of *multiscale analysis* (Mallat [1987], Meyer [1986]) and which is in fact closely related to the subject of wavelet transforms.

Returning to the construction of a discrete frame, Parseval's formula gives

$$\begin{aligned} f_{mn} &= \langle\, h_{mn} \,|\, f \,\rangle = \langle\, \hat{h}_{mn} \,|\, \hat{f} \,\rangle \\ &= \int_{-\infty}^{\infty} d\nu\; \overline{k_{mn}(\nu)}\, \hat{f}(\nu) \end{aligned} \tag{28}$$

where $k_{mn}(\nu) \equiv \hat{h}_{mn}(\nu)$ is the Fourier transform of $h_{mn}(t)$,

$$k_{mn}(v) = a^{m/2} \exp(2\pi i n a^m b \nu)\, \hat{h}(a^m \nu). \tag{29}$$

We now assume that $k(\nu) \equiv \hat{h}(\nu)$ vanishes outside the interval

$$I_0 = \{F/a \le \nu \le Fa\}, \tag{30}$$

where $F > 0$ is some fixed frequency to be determined below. The width of the "band" I_0 is $W_0 = (a - a^{-1})F$. Therefore the function $\overline{k(a^m\nu)}\hat{f}(\nu)$ is supported on the compact interval

$$I_m = \left[F/a^m,\, F/a^{m-1}\right] \tag{31}$$

of width $W_m = W_0/a^m$, and we can expand it in a Fourier series in that interval:

$$\overline{k(a^m\nu)}\, \hat{f}(\nu) = \sum_n \exp(2\pi i n \nu / W_m)\, c_{mn}, \tag{32}$$

where

$$c_{mn} = W_m^{-1} \int_{-\infty}^{\infty} d\nu\, \exp(-2\pi i n \nu / W_m)\, \overline{k(a^m\nu)}\, \hat{f}(\nu). \tag{33}$$

Comparing this with

$$f_{mn} = \int_{-\infty}^{\infty} d\nu \, a^{m/2} \, \exp(-2\pi i \nu n b a^m) \, \overline{k(a^m \nu)} \hat{f}(\nu) \tag{34}$$

suggests that we choose F so that $W_m = 1/a^m b$, which gives

$$F = \frac{a}{(a^2 - 1)\, b} \tag{35}$$

and

$$c_{mn} = a^{m/2} b f_{mn}. \tag{36}$$

The Fourier series representation above only holds in the interval I_m, since the left–hand side vanishes outside this interval while the right–hand side is periodic. To get equality for all frequencies and reconstruct $f(t)$, multiply both sides by $k(a^m \nu)$ and sum over m:

$$\begin{aligned} \left(\sum_m | \, k(a^m \nu) \, |^{\, 2} \right) \, \hat{f}(\nu) &= b \sum_{m,n} a^{m/2} \exp\left(2\pi i n a^m b \nu\right) \, k(a^m \nu) \, f_{mn} \\ &= b \sum_{m,n} k_{mn}(\nu) \, f_{mn}. \end{aligned} \tag{37}$$

To have a frame we would need the series on the left–hand side to converge to a function $\chi^+(\nu)$ with

$$0 < A \leq \chi^+(\nu) \equiv \sum_m | \, k(a^m \nu) \, |^{\, 2} \leq B \tag{38}$$

for some constants A and B. But this is a priori impossible, since $\hat{h}(\nu)$ is supported on an interval of positive frequencies and $a^m > 0$, so $\chi^+(\nu)$ vanishes for $\nu \leq 0$. However, we can choose $h(t)$, a and b such that $\chi^+(\nu)$ satisfies the frame condition for $\nu > 0$. Negative frequencies will be taken care of by starting with the complex–conjugate of the original wavelet. We adopt the notation

$$h^+(t) \equiv h(t), \quad h^-(t) \equiv \overline{h(t)}. \tag{39}$$

Then the Fourier transforms $k^\pm$ of $h^\pm$ are related by

$$k^-(\nu) = \overline{k^+(-\nu)}, \tag{40}$$

hence $k^-(\nu)$ is supported on $-I_0 = [-Fa,\, -F/a]$. A similar argument to the above gives

$$\left(\sum_m |\, k^-(a^m \nu)\,|^{\,2} \right) \hat{f}(\nu) = b \sum_{m,n} k^-_{mn}(\nu)\, f_{mn} \tag{41}$$

and

$$\chi^-(\nu) \equiv \sum_m |\, k^-(a^m \nu)\,|^{\,2} = \chi^+(-\nu). \tag{42}$$

Hence if χ^+ satisfies the frame condition for $\nu > 0$, then

$$0 < A \le \chi^+(\nu) + \chi^-(\nu) \le B \tag{43}$$

for all $\nu \neq 0$. Since $\{0\}$ has zero measure in frequency space, the frame condition is satisfied by the joint set of vectors

$$\mathcal{H}^{a,b}_M = \{k^+_{mn},\, k^-_{mn} \,|\, m, n \in \mathbb{Z}\}. \tag{44}$$

The metric operator

$$G \equiv \sum_{\epsilon=\pm} \sum_{m,n\in\mathbb{Z}} |\, k^\epsilon_{mn}\,\rangle\langle\, k^\epsilon_{mn}\,| \tag{45}$$

is given by

$$(Gf)(t) = \left((\chi^+ + \chi^-)\hat{f} \right)\check{}\,(t) \tag{46}$$

and satisfies the frame condition $0 < AI \le G \le B$. Since G is no longer a multiplication operator in the time domain (as was the case with the discrete frame connected to the windowed Fourier transform), the recovery of signals would be greatly simplified if the frame was tight. The following construction is borrowed from Daubechies [1988a]. Let $F = a/(a^2-1)b$ as above and let k be any non–negative integer or $k = \infty$. Choose a real–valued function $\eta \in C^k(\mathbb{R})$ (i.e., η is k times continuously differentiable) such that

$$\eta(x) = \begin{cases} 0 & \text{for } x \le 0 \\ \pi/2 & \text{for } x \ge 1. \end{cases} \tag{47}$$

(Such functions are easily constructed; they are used in differential geometry, for example, to make partitions of unity; see Warner [1971].) Define $h(t)$ through its Fourier transform $k^+(\nu)$ by

$$k^+(\nu) = \begin{cases} \sin\left[\eta\left(\frac{\nu - F/a}{F - F/a}\right)\right] & \text{for } \nu \le F \\ \cos\left[\eta\left(\frac{\nu - F}{aF - F}\right)\right], & \text{for } \nu \ge F. \end{cases} \tag{48}$$

Note that $k^+(\nu)$ is C^k since the derivatives of $\eta(x)$ up to order k all vanish at $x = 0$ and $x = \pi/2$. This means that the wavelets in the frame we are about to construct are all C^k. Also, k^+ vanishes outside the interval $I_0 = [F/a, Fa]$. The width of its support is $W_0 = (a - a^{-1})F$, and for each frequency $\nu > 0$ there is a unique integer M such that $F/a < a^M\nu \le F$, hence also $F < a^{M+1}\nu \le aF$. Therefore, for $\nu > 0$,

$$\chi^+(\nu) = \sin^2\left[\eta\left(\frac{\nu - a^{-1}F}{F - a^{-1}F}\right)\right] + \cos^2\left[\eta\left(\frac{\nu - F}{aF - F}\right)\right] = 1. \tag{49}$$

Thus

$$\chi^+(\nu) = \begin{cases} 0 & \text{for } \nu \le 0 \\ 1 & \text{for} \nu > 0, \end{cases} \tag{50}$$

i.e., $\chi^+(\nu)$ is the indicator function for the set of positive numbers. It follows that $\chi^-(\nu)$ is the indicator function for the negative reals, and

$$\chi^+ + \chi^- = 1 \quad \text{a.e.} \tag{51}$$

This choice of k^+ and $k^- = \overline{k^+}$ gives us a tight frame,

$$\sum_{\epsilon=\pm} \sum_{m,n\in\mathbb{Z}} |\, k^\epsilon_{mn} \,\rangle\langle\, k^\epsilon_{mn} \,| = I. \tag{52}$$

This frame is not a basis; if it were, it would have to be an orthonormal basis since it is a normal frame, hence the reproducing kernel would have to be diagonal. But

$$K(\epsilon, m, n;\ \epsilon', m', n') \equiv \langle\, k^\epsilon_{mn} \,|\, k^{\epsilon'}_{m'n'} \,\rangle \tag{53}$$

does not vanish for $\epsilon' = \epsilon$, $n' = n$ and $m' = m \pm 1$, due to the overlap of wavelets with adjacent scales. However, it is possible to construct orthonormal bases of wavelets which, in addition, have some other

surprising and remarkable properties. For example, such bases have been found (Meyer [1985], Lemarie and Meyer [1986]) whose Fourier transforms, like those above, are C^∞ with compact support and which are, simultaneously, unconditional bases for all the spaces $L^p(\mathbb{R})$ with $1 < p < \infty$ as well as all the Sobolev spaces and some other popular spaces to boot. Similar bases were constructed in connection with quantum field theory (Battle [1987]) which are only C^k for finite k but, in return, are better localized in the time domain (they have exponential decay). The concept of multiscale analysis (Mallat [1987], Meyer [1986]) provided a general method for the construction and study of orthonormal bases of wavelets. This was then used by Daubechies [1988b] to construct orthonormal bases of wavelets having *compact* support and arbitrarily high regularity.

The mere existence of such bases has surprised analysts and made wavelets a hot new topic in current mathematical research. They are also finding important applications in a variety of areas such as signal analysis, computer science and quantum field theory. They are the subject of the next chapter, where a new, algebraic, method is developed for their study.

Chapter 3
FRAMES AND LIE GROUPS

3.1. Introduction

Although we have not sought to exploit it until now, it is clear that all our frames so far have been obtained with the aid of group operations. The frames associated with the canonical coherent states and the windowed Fourier transform were built using translations in phase space (Weyl–Heisenberg group), while the wavelet frames used translations and dilations (the affine group). In this chapter we look for a unifying pattern in these constructions based on group theory. We analyze the foregoing constructions in turn, and draw separate lessons from each. It will be natural to work in reverse order. The affine group, which is, in some sense, the simplest, will lead us to the general method. Successive refinements will be suggested by the windowed Fourier transform and the canonical coherent states.

3.2. Klauder's Group–Frames

This was the first of the group–theoretic constructions, pioneered by J. R. Klauder [1960, 1963a, 1963b], who was also the first to apply it to the affine group $G = \mathbb{R}^* \times \mathbb{R}$ (Aslaksen and Klauder [1968, 1969]). An element $g = (a, s)$ of G acts on the real line ("time") by dilation followed by translation:

$$gt \equiv (a, s)t = at + s. \tag{1}$$

The group–composition law is given by

$$g'gt = (a', s')(a, s)t = a'(at + s) + s' = (a'a, a's + s')t, \tag{2}$$

hence $g^{-1} = (a^{-1}, -s/a)$. (The form of the composition law shows that the subgroup of dilations *acts* on the subgroup of translations,

so that G is the semidirect product of the two.) The frame vectors for the wavelet transform were obtained from a single "basic wavelet" vector h by applying the transformation

$$h(t) \mapsto \, \mid a \mid^{-1/2} h\left(\frac{t-s}{a}\right) \equiv (U(a,s)h)\,(t). \tag{3}$$

It it easy to see that with respect to the inner product in $L^2(\mathbb{R})$,
(a) $(U(a,s)^*h)\,(t) = \mid a \mid^{1/2} h(at+s) = \left(U(a,s)^{-1}h\right)(t)$, and
(b) $U(a',s')U(a,s)h = U(a'a, a's+s')h$.
In terms of the group operations, this means that $U(g)^* = U(g)^{-1}$ and $U(g'g) = U(g')U(g)$, respectively. That is, each $U(g)$ is a *unitary* operator (on $L^2(\mathbb{R})$), and $g \mapsto U(g)$ is a *representation* of G. This means that U is a *unitary representation* of G on $L^2(\mathbb{R})$. The theory of such representations for general groups is a deep and highly developed subject, and is of fundamental importance in quantum mechanics, as was realized by Hermann Weyl and others long ago (Weyl [1931]). In the broadest sense, group representations amount to a vast generalization of the exponential function (think of the map $z \mapsto e^{az}$ from $\mathbb{C}$ to $\mathbb{C}^*$), and unitary group representations generalize the map $x \mapsto e^{iax}$ from $\mathbb{R}$ to the unit circle.

A representation U of a group G on a Hilbert space $\mathcal{H}$ is said to be *reducible* if $\mathcal{H}$ has a non–trivial closed subspace S (i.e., $S \neq \{0\}$ and $S \neq \mathcal{H}$) which is *invariant* under U (i.e., $U(g)S \subseteq S$ for every $g \in G$). If U is unitary and S is invariant, then clearly so is its orthogonal complement $S^{\perp}$. If no such S exists, then U is said to be *irreducible*. It can be shown that the above representation of $\mathbb{R}^* \times \mathbb{R}$ on $L^2(\mathbb{R})$ is, in fact, irreducible.

The method to be described below assumes that we begin with a given irreducible unitary representation U of a given group G on a given Hilbert space $\mathcal{H}$. In addition, we assume that the group is a *Lie group*, meaning that it has a differetiable structure such that it is essentially determined (in local terms) by its Lie algebra of left– (or right–) invariant vector fields. (See Helgason [1978], Varadarajan [1974] or Warner [1971] for background on Lie groups.) The affine group and the Weyl–Heisenberg group are examples of Lie groups, as is $\mathbb{R}^n$ (under vector addition as the group operation).

Given a general setup (U, G, $\mathcal{H}$) as above, choose an arbitrary non–zero vector h in $\mathcal{H}$. (Klauder dubbed h a "fiducial vector"; for the affine group, this was the "basic wavelet".) For every $g \in G$ define the vector

$$h_g = U(g)h. \tag{4}$$

Since U is unitary, $\|h_g\| = \|h\|$. The h_g's are *covariant* under the action of G on $\mathcal{H}$, i.e.,

$$U(g')h_g = U(g')U(g)h = U(g'g)h = h_{g'g}. \tag{5}$$

Hence the set of all finite linear combinations (span) of h_g's is invariant under U, and therefore so is its closure S. Since U is irreducible and $S \neq \{0\}$, it follows that $S = \mathcal{H}$. This means that every vector in $\mathcal{H}$ can be approximated to arbitrary precision by finite linear combinations of h_g's—a good beginning, if one is ultimately interested in reconstruction! To build a frame from the h_g's, we need a measure on G. Now every Lie group has an essentially unique (up to a constant factor) *left–invariant* measure, which we will denote by $d\mu$. This means that if E is an arbitrary (Borel) subset of G, and if $g_1E \equiv \{g_1g \,|\, g \in E\}$ is its *left translate* by $g_1 \in G$, then

$$\mu(g_1E) \equiv \int_{g_1E} d\mu(g) = \mu(E) \equiv \int_E d\mu(g). \tag{6}$$

In local terms, $d\mu(g_1g) = d\mu(g)$ for fixed g_1. [Similarly, there exists a *right–invariant* measure $d\mu_R$ on G, which is in general different from $d\mu$; if $d\mu_R$ is proportional by a constant to $d\mu$, the group G is called *unimodular*. Everything we do below can be repeated, with obvious modifications, using $d\mu_R$, provided that h_g is redefined as $U(g^{-1})h$.]

To find $d\mu$ for the affine group, for example, we write it in the form of a density,

$$d\mu(a,s) = \rho(a,s)\, da \wedge ds, \tag{7}$$

where $da \wedge ds$ is a differential 2–form denoting the area element in $\mathbb{R}^2 \supset G$. ($da \wedge ds$ becomes a positive measure upon choosing an *orientation* in $\mathbb{R}^2$ and orienting all subsets accordingly; see Warner [1971].) Then, for fixed $(a_1, s_1) \in \mathbb{R}^* \times \mathbb{R}$,

$$\begin{aligned} d\mu((a_1,s_1)(a,s)) &= \rho(a_1a, a_1s+s_1)\, d(a_1a) \wedge d(a_1s+s_1) \\ &= a_1^2 \rho(a_1a, a_1s+s_1)\, da \wedge ds, \end{aligned} \tag{8}$$

hence left–invariance implies

$$a_1^2 \rho(a_1 a, a_1 s + s_1) = \rho(a, s). \tag{9}$$

Setting $a_1 = 1/a$ and $s_1 = -s/a$ then shows that

$$d\mu(a, s) = da \wedge ds/a^2, \tag{10}$$

where we have chosen the normalization $\rho(1, 0) = 1$. This is precisely the measure we obtained earlier using a more pedestrian approach.

Returning to the general case, consider the *formal* integral

$$J = \int_G d\mu(g) \, | \, h_g \, \rangle \langle \, h_g \, | \, . \tag{11}$$

We want to show that (a) J converges, in some sense, and (b) J is a multiple of the identity, thus giving a tight frame in the generalized sense defined in chapter 1. Before worrying about convergence, let us formally apply $U(g_1)$ from the left:

$$\begin{aligned} U(g_1)J &= \int_G d\mu(g) \, U(g_1) \, | \, h_g \, \rangle \langle \, h_g \, | \\ &= \int_G d\mu(g) \, | \, h_{g_1 g} \, \rangle \langle \, h_g \, | \\ &= \int_G d\mu(g') \, | \, h_{g'} \, \rangle \langle \, h_{g_1^{-1} g'} \, | \\ &= \int_G d\mu(g') \, | \, h_{g'} \, \rangle \langle \, h_{g'} \, | \, U(g_1) \\ &= JU(g_1), \end{aligned} \tag{12}$$

where we have used the left–invariance of the measure and

$$\langle \, h_{g_1^{-1} g'} \, | \, = \Big(| \, h_{g_1^{-1} g'} \, \rangle \Big)^* = \big(U(g_1^{-1}) \, | \, h_{g'} \, \rangle \big)^* = \langle \, h_{g'} \, | \, U(g_1), \tag{13}$$

which follows from unitarity.

That is, J commutes with every representative $U(g)$ of the group. By Schur's Lemma (Varadarajan [1974]), it follows from the irreducibility of U that J is a multiple of the identity operator on $\mathcal{H}$. Since J, if it converges, is a positive operator, we arrive at the desired frame condition

$$\int_G d\mu(g)\,|\,h_g\,\rangle\langle\,h_g\,|\,=cI, \tag{14}$$

where c is a positive constant which can taken as unity by the appropriate normalization of $d\mu$. Returning to the formal integral defining J, the above argument shows that *if* the integral converges in some sense, it must converge to cI, hence define a bounded operator. (Of course c could be infinite!) Thus a *necessary* condition for convergence in the *weak* sense (i.e., as a quadratic form) is that

$$\begin{aligned}\langle\,h\,|\,J\,h\,\rangle &= \int_G d\mu(g)\,\langle\,h\,|\,h_g\,\rangle\langle\,h_g\,|\,h\,\rangle \\ &= \int_G d\mu(g)\;|\,\langle\,h\,|\,U(g)\,|\,h\,\rangle\,|^{\,2}<\infty.\end{aligned} \tag{15}$$

We have already encountered this condition in the special case of the affine group, where it was called the *admissibility condition* for h. The same terminology is used in the present, general setting. The above condition depends both on the representation U and the choice of h. If it is satisfied for at least *one* non–zero vector h, the representation U is called *square–integrable* and h is called *admissible*. It turns out that the existence of an admissible (non–zero) vector is also *sufficient* for the weak convergence of the integral J. The following theorem is due to Carey [1976], and Dufflo and Moore [1976]; G can be any locally compact topological group, in particular any Lie group.

Theorem 3.1. *Let U be a square–integrable unitary irreducible representation of G on a Hilbert space $\mathcal{H}$. Then there exists a unique self–adjoint (in general unbounded) operator C on $\mathcal{H}$ such that:*
(a) The domain of C coincides with the set of all admissible vectors.
(b) If h_1 and h_2 are admissible, then for all f_1 and f_2 in $\mathcal{H}$ we have

$$\int_G d\mu(g)\,\langle\,f_1\,|\,U(g)\,h_1\,\rangle\langle\,U(g)\,h_2\,|\,f_2\,\rangle=\langle\,Ch_2\,|\,Ch_1\,\rangle\langle\,f_1\,|\,f_2\,\rangle. \tag{16}$$

(c) If G is unimodular, then C is a multiple of the identity.

Choosing $h_1 = h_2 \equiv h$ now leads to the earlier resolution of unity, with $c = \|Ch\|^2$. As usual, an arbitrary vector f in $\mathcal{H}$ can be "presented" as a function $\tilde{f}(g) \equiv \langle\,h_g\,|\,f\,\rangle$ on G, from which f may be reconstructed as a linear combination of h_g's weighted by the reproducing kernel $K(g',g)=\langle\,h_{g'}\,|\,h_g\,\rangle$.

Note that the basic wavelet h corresponds to

$$\tilde{h}(g) = \langle\, h_g \,|\, h \,\rangle = \overline{\langle\, h \,|\, U(g)h \,\rangle}, \tag{17}$$

so the admissibility condition is nothing but the requirement that $\tilde{h}$ belong to $L^2(d\mu)$.

A potentially interesting generalization of this scheme is actually possible. By the foregoing theorem we can use two *distinct* admissible vectors: the vector h_2 is used to *analyze* f, i.e. present it as $\tilde{f}(g) = \langle\, (h_2)_g \,|\, f \,\rangle$, while the vector h_1 is then used to *synthesize* f from $\tilde{f}(g)$. There is no fundamental reason to use the same "wavelets" for analysis and synthesis, provided only that they overlap in the sense that

$$\langle\, Ch_2 \,|\, Ch_1 \,\rangle \neq 0. \tag{18}$$

Absorbing the reciprocal of this constant into the group measure, the map T: $f \mapsto \tilde{f}$ is an isometry from $\mathcal{H}$ onto its range $\Re_T$.

Due to the "covariance" of the h_g's with respect to the action of G, the representation of G on $\Re_T$ acquires the simple *geometric* form

$$\begin{aligned} \left(\tilde{U}(g_1)\tilde{f}\right)(g) &\equiv \langle\, h_g \,|\, U(g_1)f \,\rangle \\ &= \langle\, h_{g_1^{-1}g} \,|\, f \,\rangle \\ &= \tilde{f}(g_1^{-1}g), \end{aligned} \tag{19}$$

showing that G acts on $\Re_T$ by merely translating the variable in the base space G. The realization of f by $\tilde{f}$ and U by $\tilde{U}$ as above is called the *coherent–state representation* determined by the pair (U, h). We will also refer to the frame $\{\, h_g \,|\, g \in G\}$ as the *group–frame* (G–frame) associated with (U, h). From a purely mathematical point of view, one of the attractions of this scheme is that although we started with an arbitrary representation of G on an arbitrary Hilbert space, this construction "brings it home" to G itself and objects directly associated with it: the Hilbert space is a closed subspace of $L^2(d\mu)$, and the representation is induced from the (left) action of G on itself, i.e. $g \mapsto g_1^{-1}g$. That is, Klauder's construction exhibits U as a subrepresentation of the *regular representation* of G. (See Mackey [1968] for the definition and discussion of the regular representation.)

3.3. Perelomov's Homogeneous G–Frames

Let us now attempt to apply the group–theoretic method to the windowed Fourier transform. As most of our applications will be to the phase–space formulation of quantum mechanics, we shift gears and replace the time by a space coordinate, $t \to -x$ (the sign is related to the Minkowski metric; see section 1.1) and the frequency by a momentum coordinate, $2\pi\nu \to p$. Although it is a trivial matter to extend everything we do in this section to an arbitrary (finite) number of degrees of freedom (where x and p belong to $\mathbb{R}^s$), we restrict ourselves to a single degree of freedom to keep the notation simple. The self–adjoint generators of translations in space and momentum, P and X, are defined by

$$Pf(x) = -i\frac{\partial}{\partial x}f(x), \qquad (Xf)\hat{}(p) = i\frac{\partial}{\partial p}\hat{f}(p). \tag{1}$$

Expressed in the space domain,

$$Xf(x) = xf(x). \tag{2}$$

Starting with a basic window function $h(x)$, the window centered at (p, x) in phase space is given by

$$\begin{aligned} h_{p,x}(x') &= e^{ipx'}h(x'-x) \\ &= e^{ipx'}\left(e^{-ixP}h\right)(x') \\ &= \left(e^{ipX}\,e^{-ixP}h\right)(x'). \end{aligned} \tag{3}$$

That is, $h_{p,x}$ is obtained from h by a translation in space (by x) followed by a translation in momentum (by p). Defining the corresponding unitary operators

$$U(p,x) = e^{ipX}\,e^{-ixP}, \tag{4}$$

let us see what happens when two such operations are applied in succession:

$$\begin{aligned} \left(U(p_1,x_1)U(p,x)h\right)(x') &= \left(U(p_1,x_1)h_{p,x}\right)(x') \\ &= e^{ip_1x'}h_{p,x}(x'-x_1) \\ &= e^{ip_1x'}e^{ip(x'-x_1)}h(x'-x_1-x) \\ &= e^{-ipx_1}h_{p_1+p,x_1+x}(x') \\ &= e^{-ipx_1}U(p_1+p,x_1+x)h(x'). \end{aligned} \tag{5}$$

Hence the operators $U(p,x)$ do not form a group, since two successive operations give rise to a "multiplier" $\exp(-ipx_1)$. The reason is that translations in space do not commute with translations in momentum, as can also be seen at the infinitesimal level by noting that their respective generators obey the "canonical commutation relations"

$$[X,P] = \left[x, -i\frac{\partial}{\partial x}\right] = iI. \tag{6}$$

The remedy (suggested by Hermann Weyl) is to include the identity operator as a new generator (it generates phase factors $e^{-i\phi}$ which can be used to absorb the multiplier). To see this in terms of unitary representations of Lie groups, consider the abstract *real* Lie algebra $\mathbf{w}$ with three generators $\{-iS, -iT, -iE\}$ and Lie brackets

$$[S,T] = iE, \qquad [S,E] = 0, \qquad [T,E] = 0. \tag{7}$$

The corresponding three–dimensional (simply connected) Lie group $\mathcal{W}$ is known as the *Weyl–Heisenberg group*. Topologically, $\mathcal{W}$ is just $\mathbb{R}^3$. (If configuration space is $\mathbb{R}^s$, the corresponding Weyl–Heisenberg group $\mathcal{W}_s \approx \mathbb{R}^{2s+1}$.) A general element in $\mathcal{W}$, parametrized by $(p,x,\phi) \in \mathbb{R}^3$, may be expressed as the product of three factors

$$g(p,x,\phi) = \exp(-i\phi E)\exp(ipS)\exp(-ixT), \tag{8}$$

where "exp" denotes the exponential mapping from the Lie algebra $\mathbf{w}$ to the Lie group $\mathcal{W}$, whose group law is

$$g(p_1,x_1,\phi_1)\, g(p,x,\phi) = g(p_1+p, x_1+x, \phi_1+\phi+px_1). \tag{9}$$

A unitary irreducible representation of $\mathcal{W}$ on $L^2(\mathbb{R})$ is obtained by the correspondence $S \to X$, $T \to P$, $E \to I$. This is known as the *Schrödinger representation*. The unitary operator correponding to $g(x,p,\phi)$ is

$$U(p,x,\phi) = e^{-i\phi}\, e^{ipX}\, e^{-ixP}. \tag{10}$$

As expected, these operators are closed under multiplication, with the composition law

$$U(p_1,x_1,\phi_1)\, U(p,x,\phi) = U(p_1+p, x_1+x, \phi_1+\phi+px_1).$$

With this unitary irreducible representation of $\mathcal{W}$ on $L^2(\mathbb{R})$, we have all the ingredients needed to attempt the construction of a group frame for $\mathcal{W}$. The prospective frame vectors are

$$h_{p,x,\phi} = U(p,x,\phi)h = e^{-i\phi}h_{p,x}, \tag{11}$$

where $h_{p,x}$ are the vectors defined earlier. The left–invariant measure on $\mathcal{W}$ (which, in this case, is also right–invariant) is just Legesgue measure on $\mathbb{R}^3$, given by the differential form $dp \wedge dx \wedge d\phi$. This can be seen by looking at the composition law: For fixed (p_1, x_1, ϕ_1),

$$d(p_1 + p) \wedge d(x_1 + x) \wedge d(\phi_1 + \phi + px_1) = dp \wedge dx \wedge d\phi \tag{12}$$

since $dp \wedge dp = 0$. The method of section 2 then gives the following candidate for a resolution of unity:

$$\begin{aligned} J &= \int_{\mathcal{W}} dp \wedge dx \wedge d\phi \,|\, h_{p,x,\phi} \,\rangle\langle\, h_{p,x,\phi} \,| \\ &= \int_{\mathcal{W}} dp \wedge dx \wedge d\phi \,|\, h_{p,x} \,\rangle\langle\, h_{p,x} \,|\,, \end{aligned} \tag{13}$$

where the phase factor cancels in the integrand. This integral clearly diverges, since the integrand is independent of ϕ and the integration is over all real ϕ. Equivalently, the representation U is not square–integrable, since for every nonzero h,

$$c_h \equiv \int dp\, dx\, d\phi \;|\, \langle\, h \,|\, U(p,x,\phi)h \,\rangle\, |\,^2 = \infty \tag{14}$$

due, again, to the constancy of the integrand in ϕ. We could get around the problem by choosing a *multiply connected* version $\mathcal{W}_1$ of $\mathcal{W}$, say with $0 \le \phi < 2\pi$. ($\mathcal{W}_1$ has the same Lie algebra as $\mathcal{W}$). This would indeed give a tight frame, but this frame is *unnecessarily* redundant (as opposed to the beneficial sort of redundance associated with oversampling) since the vectors $h_{p,x,\phi}$ are not essentially different for distinct values of ϕ. More significantly, we would miss an important lesson which this example promises to teach us. For other important groups, as we will see, the problem cannot be circumvented by compactifying the troublesome parameters. The following solution was proposed by Perelomov [1972] (see also Klauder [1963b, p. 1068], where this idea is anticipated). To get rid of the ϕ–dependence, choose a "slice" of $\mathcal{W}$, $\phi = \alpha(p,x)$, and define

$$h^{\alpha}_{p,x} = h_{p,x,\alpha(p,x)} = e^{-i\alpha(p,x)} h_{p,x}. \tag{15}$$

Integrating only over this slice with the measure $dp \wedge dx$, we get

$$J' \equiv \int dp \wedge dx \; | \, h^{\alpha}_{p,x} \, \rangle \langle \, h^{\alpha}_{p,x} \, | \; = c_h I, \tag{16}$$

where $c_h = 2\pi \|h\|^2$, since the integral reduces to the same one we had for the windowed Fourier transform.

From a computational point of view this is, of course, trivial. But to extend the technique to other groups we must understand the group theory behind it. Suppose, then, that we return to the general setup we had in the last section: Given a unitary irreducible representation U of a Lie group G on a Hilbert space $\mathcal{H}$, choose a nonzero vector h in $\mathcal{H}$ and form its translates $h_g = U(g)h$ under the group action as before. Consider now the set H of of all elements k of G for which the action of $U(k)$ on h reduces to a multiplication by a phase factor $\chi(k) = \exp[i\phi(k)]$:

$$U(k)h = \chi(k)h, \qquad k \in H. \tag{17}$$

In the case of $\mathcal{W}$, H consists of all elements of the form $k = (0, 0, \phi)$ and $U(k)$ is, in fact, nothing but a phase factor. However, this is deceptive, since in general $U(k)$ may be a non–trivial operator, acting trivially only on some vectors h. Hence, in general, H will in fact depend on the choice of h and should properly be designated $H(h)$. Now for two elements k_1 and k_2 of H, we have

$$U(k_1 k_2^{-1})h = U(k_1)U(k_2)^{-1}h = \chi(k_1)\chi(k_2)^{-1}h, \tag{18}$$

hence $k_1 k_2^{-1}$ also belongs to H and it follows that H is a *subgroup* of G. Furthermore, the above equation shows that the map $k \mapsto \chi(k)$ is a group–homomorphism of H into the unit circle, thus it is a *character* of H, i.e. a unitary representation on the one–dimensional Hilbert space $\mathbb{C}$. H is called the *stability subgroup* for h, and its Lie algebra is called the stability subalgebra. The reason for this terminology is that H does not affect the quantum–mechanical *state* defined by h, since all observable expectation values in that state are given in terms of sequilinear forms in h. More precisely, if we choose $\|h\| = 1$, the corresponding state is by definition the rank–one projection operator

$$P = \; | \, h \, \rangle \langle \, h \, | \tag{19}$$

and the expected value of any observable A in this state is given by

$$\langle A \rangle \equiv \langle h \,|\, A\, h \rangle = \text{trace}\,(PA). \tag{20}$$

(This formulation, besides avoiding irrelevant phase facors, also permits states which are statistical mixtures of pure states as needed, for example, in statistical quantum mechanics.) The translate of P under a general group element g is then

$$P_g \equiv |\, h_g \,\rangle\langle\, h_g \,| = U(g)\, P\, U(g)^*, \tag{21}$$

hence for g in H we have $P_g = P$, i.e., P is *stable* under H as the name implies. There is, therefore, some advantage to formulating the theory as much as possible in terms of states rather than Hilbert space vectors since this automatically eliminates the fictitious degree of freedom represented by the overall phase. [Incidentally, a similar situation appears to prevail in communication theory, since an overall phase shift has no effect whatsoever on the informational content of the signal.] However, the Hilbert space vectors will play an important role in connection with holomorphy (recall that in the coherent–state representation, $\tilde{f}(z)$ is analytic, whereas $|\,\tilde{f}(z)\,|^{\,2}$ is not), hence we work primarily with them.

If $g \in G$ and $k \in H$, then

$$h_{gk} = U(gk)h = U(g)U(k)h = \chi(k)h_g, \tag{22}$$

hence $P_{gk} = P_g$. That is, P_g is the same for all members of the *left coset*

$$gH \equiv \{gk \,|\, k \in H\}. \tag{23}$$

The set of all translates P_g is therefore parametrized by the *left coset space*

$$M = G/H \equiv \{gH \,|\, g \in G\}. \tag{24}$$

Members of M will be alternatively denoted by m and by gH, and will play a dual role: as points in M, and subsets of G. In the case of $\mathcal{W}$, for example, M is parametrized by $(p, x) \in \mathbb{R}^2$, i.e. it is *phase space*, precisely the label space for the frame we obtained. The coset $(p, x)H$ is a straight line in $\mathcal{W} \approx \mathbb{R}^3$. (We will see in the next chapter that $\mathcal{W}$ can be interpreted as a degenerate non–relativistic limit of phase space×time and $(p, x)H$ then corresponds to the trajectory of a free classical particle in $\mathcal{W}$.)

To build a frame, we take a "slice" of G by choosing a representative from each coset, i.e. choosing a map

$$\sigma\colon G/H \to G. \tag{25}$$

Note: This is actually a non–trivial process. The projection $G \to G/H$ defines a *fiber bundle*, and σ is a *section* of this bundle; in general σ can be chosen *smoothly* only in a neighborhood of each point of G/H, i.e., locally (see F. Warner [1971], p.120). However, this is sufficient for our purposes, since we will ultimately deal only with the state P_m, which is independent of the choice of σ. #

Thus σ has the form $\sigma(gH) = g\,\alpha(g)$ for some function $\alpha\colon G \to H$. In the case of $\mathcal{W}$, we had $\sigma(p,x) = (p,x,\alpha(p,x))$. The frame vectors corresponding to this choice are

$$h^\sigma_m \equiv h_{\sigma(m)} = U(\sigma(m))h. \tag{26}$$

To build a frame, we need two more ingredients: an *action* of G on the label space M, and a measure on M which is invariant with respect to this action. The action is easy, since G acts naturally on G/H by left translation:

$$g_1(gH) = (g_1 g)H. \tag{27}$$

If $m = gH \in M$, we will denote $(g_1 g)H$ by $g_1 m$. M is called a *homogeneous space* of G. As for an invariant measure, it exists, in general, only subject to a certain technical condition (Helgasson [1962], p. 369). It does exist whenever G is a unimodular group (its right– and left–invariant measures are proportional by a constant), such as $\mathcal{W}$. Let us assume that a G–invariant measure does exist on M (in which case it is unique, up to a constant factor) and denote it by $d\mu_{_M}$. Once more we consider the formal integral

$$J = \int_M d\mu_{_M}(m)\,|\,h^\sigma_m\,\rangle\langle\,h^\sigma_m\,| \;= \int_M d\mu_{_M}(m)\,P_m. \tag{28}$$

If we can show that J commutes with every $U(g_1)$, then irreducibility again forces $J = cI$. But

$$\begin{aligned} U(g_1)\, h^\sigma_{gH} &= U(g_1)\, U(\sigma(gH))h \\ &= U(g_1\sigma(gH))\, h \\ &= U(g_1 g\, \alpha(g))\, h \\ &= U(g_1 g\, \alpha(g_1 g)\alpha(g_1 g)^{-1}\alpha(g))\, h \\ &= \chi(\alpha(g_1 g)^{-1}\alpha(g))\, h^\sigma_{g_1 gH}. \end{aligned} \tag{29}$$

That is, the vectors h^σ_m are "almost" covariant under the action of G, with a residual phase factor (multiplier). This means that the *states* P_m transform covariantly, i.e.,

$$U(g_1)\, P_m\, U(g_1)^* = P_{g_1 m}. \tag{30}$$

Thus

$$\begin{aligned} U(g_1)\, J\, U(g_1)^{-1} &= \int_M d\mu_{_M}(m)\, U(g_1)\, P_m\, U(g_1)^* \\ &= \int_M d\mu_{_M}(m)\, P_{g_1 m} \\ &= \int_M d\mu_{_M}(m')\, P_{m'} \\ &= J, \end{aligned} \tag{31}$$

using the unitarity of $U(g_1)$ and the invariance of $d\mu_{_M}$. This proves that $J = cI$, provided the integral converges in some sense.

Note: The above proof becomes shorter if one works directly with the states P_m rather than the frame vectors; however, it is important to see the action of G on the frame vectors in terms of the multipliers since this will play a role in the next section, where we look for representations of G on spaces of holomorphic functions. #

A necessary condition for the *weak* convergence of the integral is that

$$\langle\, h \,|\, J\, h \,\rangle = \int_M d\mu_{_M}(m)\; |\, \langle\, h^\sigma_m \,|\, h \,\rangle\, |^{\,2} < \infty, \tag{32}$$

i.e., that $\tilde{h}(m) \equiv \langle\, h^\sigma_m \,|\, h \,\rangle$ be in $L^2(d\mu_{_M})$. As before, this condition is also sufficient, and the same terminology is used: h is called admissible, and U is called square–integrable. The action of U in the

Hilbert space of functions $\tilde{f}(m) = \langle\, h^{\sigma}_{m} \,|\, f \,\rangle$ is somewhat more complicated than earlier because of the multiplier. (If we simply dropped the multipliers we would no longer have a unitary representation of G but a *projective* representation; see Varadarajan [1970].)

The above construction has the advantage that the trivial part of the action of G is factored out, thereby improving the chances that the integral J converges. As mentioned above, it depends on the existence of the invariant measure $d\mu_{_M}$ on G/H, which is not guaranteed. Note that for fixed $g \in G$, the stability subgroup for the vector $U(g)h$ is gHg^{-1}, i.e., a subgroup of G *conjugate* to H. In general, however, the stability subgroups of two admissible vectors h_1 and h_2 may not be conjugates. It may happen that G/H_1 has an invariant measure while G/H_2 does not. It pays, therefore, to choose the vector h very carefully. Intuition suggests that h should be chosen so as to *maximize* the stability subgroup, since this will minimize the homogeneous space M and improve the chances for convergence. Furthermore, maximal use of symmetry would seem to make it more likely that an invariant measure exists on the quotient. More will be said about this in the next section, in connection with the "weight" of a representation.

We will refer to the frame $\{\, h^{\sigma}_{m} \,\}$ as a *homogeneous G–frame* associated with (U, h). The dependence on σ will usually be suppressed, since a change in (the local sections) σ gives an equivalent frame.

3.4. Onofri's Holomorphic G–Frames

The frames associated with the windowed Fourier transform and the canonical coherent states are similar in that both provide "phase space" representations of functions in $L^2(\mathbb{R}^s)$. The distinguishing feature is that the representation determined by the canonical coherent states is in terms of *holomorphic* (analytic) functions. Neither of the two general group–theoretic methods covered so far explains the origin of this analyticity, yet it has been found that all such group–related phase–space representations have analytic counterparts. Moreover, analyticity will play a key role in the coherent–state representations of relativistic quantum mechanics (next chapter) and quantum field theory (chapter 5). The method to be described in this section assumes we are dealing with a compact, semisimple group. Since the coherent–state representations we will develop for relativistic quantum mechanics are based on the Poincaré group, which is neither compact nor semisimple, the present considerations do not apply di-

rectly to the main body of this book (chapters 4 and 5). Nevertheless, we describe them in considerable detail in the hope that they may shed some light on our later constructions, which are still not well–understood in general terms.

Let us return to the canonical coherent states in order to isolate the property leading to analyticicy and find its generalization to other groups. We follow an approach first advocated by Onofri [1975]. For related developments, also see Perelomov [1986]. Again consider the three–dimensional Weyl–Heisenberg group $\mathcal{W}$, whose (real) Lie algebra $\mathbf{w}$ has a basis $\{-iS, -iT, -iE\}$ which is represented on $L^2(\mathbb{R})$ by $S \to X$, $T \to P$ and $E \to I$. Recall that we arrived at the canonical coherent states χ_z as eigenvectors of the non–hermitian operator $A = X + iP$, which represents the complex combination $S + iT$ of generators in $\mathbf{w}$. Since $\mathbf{w}$ is a real Lie algebra, we must *complexify* it in order to consider such combinations of its generators. This is done in the same way as complexifying a real vector space, namely by taking the tensor product with the field of complex numbers. With obvious notation, $\mathbf{w}_c = \mathbf{w} \otimes \mathbb{C} = \mathbf{w} + i\mathbf{w}$. As will be seen below, complex combinations such as $S \pm iT$ play a very important role in the theory of *real* Lie groups, exactly for the same reason that complex eigenvectors and eigenvalues are necessary in order to study real matrices.

We begin by rederiving the canonical coherent states from an algebraic point of view which shows their relation to the vectors $h_{p,x}$ associated with the windowed Fourier transform and pinpoints the property which makes them analytic. In the last section we found that

$$h_{p,x} = e^{ipX}\, e^{-ixP}\, h. \tag{1}$$

Now if B and C are operators such that $[B, C]$ commutes with both B and C, then the Baker–Campbell–Hausdorff formula (Varadarajan [1974]) reduces to

$$e^B e^C = e^{\frac{1}{2}[B,C]}\, e^{B+C}. \tag{2}$$

Since $[ipX, -ixP] = ipxI$, we therefore have

$$h_{p,x} = \exp(ipx/2)\, \exp\left(ipX - ixP\right)\, h. \tag{3}$$

Substituting

$$X = (A^* + A)/2 \qquad \text{and} \qquad -iP = (A^* - A)/2 \tag{4}$$

and defining $z \equiv x - ip$, we get

$$h_{p,x} = \exp(ipx/2)\, \exp\left[\bar{z}A^*/2 - zA/2\right]\, h. \tag{5}$$

So far, all our manipulations have been justifiable since we have only exponentiated skew–adoint operators. The next one is more delicate since the operators to be exponentiated are not skew–adjoint; it will be justified later. Using the Baker–Campbell–Hausdorff formula again with $B = \bar{z}A^*/2$ and $C = -zA/2$, write

$$h_{p,x} = \exp\left(ipx/2 - \bar{z}z/4\right)\, \exp\left(\bar{z}A^*/2\right)\, \exp\left(-zA/2\right)\, h. \tag{6}$$

If we choose

$$h(x') = N \exp\left(-x'^{\,2}/2\right) = \chi_0(x') \tag{7}$$

(where $N = (2\pi)^{-1/4}$ and χ_0 is just the canonical coherent state χ_z with $z = 0$), then $Ah = 0$, hence $\exp\left(-zA/2\right)\, h = h$ and

$$h_{p,x} = \exp\left(ipx/2 - \bar{z}z/4\right)\, \exp\left(\bar{z}A^*/2\right)\, \chi_0. \tag{8}$$

We claim that

$$\exp\left(\bar{z}A^*/2\right)\, \chi_0 = \chi_z. \tag{9}$$

This can be seen by applying A to the left–hand side, then using $A\chi_0 = 0$ and $[A, A^*] = 2I$:

$$\begin{aligned} A \exp\left(\bar{z}A^*/2\right)\, \chi_0 &= \left[A, \exp\left(\bar{z}A^*/2\right)\right]\, \chi_0 \\ &= \bar{z}\, \exp\left(\bar{z}A^*/2\right)\, \chi_0, \end{aligned} \tag{10}$$

which shows that $\exp\left(\bar{z}A^*/2\right)\, \chi_0$ equals χ_z up to a constant factor, which is easily shown to be unity. That is,

$$h_{p,x} = \exp\left(ipx - \bar{z}z/4\right)\, \chi_z, \tag{11}$$

so the canonical coherent states are a special case (modulo the z–dependent factor in front) of the frame vectors $h_{p,x}$ for the choice $h = \chi_0$. Note that the corresponding *states* are related by

$$|\, h_{p,x}\,\rangle\langle\, h_{p,x}\,| \;= \exp\left(-|z|^2/2\right)\, |\, \chi_z\,\rangle\langle\, \chi_z\,|\,, \tag{12}$$

giving the weight function on the right–hand side in the bargain. The reason for this is that the $h_{p,x}$'s were obtained by unitarily translating h, whereas the operator $\exp(\bar{z}A^*/2)$ which "translates" χ_0 to χ_z is not unitary but results in $\|\chi_z\| = \exp\left(|z|^2/4\right)\|\chi_0\|$; the weight function then corrects for this.

We can now justify the fine point we glossed over earlier. The expression

$$\exp(\bar{z}A^*/2)\,\exp(-zA/2)\,\chi_0 \tag{13}$$

makes sense because
(a)

$$e^{-zA/2}\chi_0 = \chi_0 \tag{14}$$

since $A\chi_0 = 0$, and
(b) χ_0 is an *analytic vector* (Nelson [1959]) for the operator A^*, since

$$\begin{aligned}\sum_{n=0}^{\infty}\frac{1}{n!}\,\|\,(\bar{z}A^*/2)^n\;\chi_0\,\| &= \sum_{n=0}^{\infty}\frac{|z|^n}{2^n\,n!}\|A^{*n}\chi_0\| \\ &= \sum_{n=0}^{\infty}\frac{|z|^n}{2^n\,n!}2^{n/2}\sqrt{n!}\,\|\chi_0\| < \infty \qquad \forall z\in\mathbb{C}\;,\end{aligned} \tag{15}$$

where $\|A^{*n}\chi_0\| = 2^{n/2}\sqrt{n!}$ follows from $[A, A^*] = 2I$ and $A\chi_0 = 0$.

As will be seen below, the key to constructing representations of real Lie groups on spaces of analytic functions will be to
(a) complexify the Lie algebra;
(b) find a counterpart to $A\chi_0 = 0$;
(c) define counterparts to $\chi_z = \exp(\bar{z}A^*/2)\,\chi_0$.

Before embarking on this task, we must make a brief excursion into the structure theory of Lie algebras. For background and details, see Helgason [1978] or Hermann [1966]. The Lie groups and –algebras we consider below are assumed to be *real* and *semisimple*. (A Lie algebra is semisimple if it contains no proper abelian ideals; a Lie group is semisimple if its Lie algebra is semisimple.) Since $\mathcal{W}$ is not semisimple (the subspace spanned by $-iE$ is an ideal of $\mathbf{w}$), these results do not apply to it, strictly speaking. However, as will be shown in section 3.6, $\mathcal{W}$ can be obtained as a (contraction) limit of a simple Lie group ($SU(2)$), and this turns out to be sufficient for our purpose. Thus, let G be an arbitrary real, semisimple Lie group and $\mathbf{g}$ its Lie algebra. The following is known:

1. Every element X of $\mathbf{g}$ defines a linear map $\mathrm{ad}\, X$ on $\mathbf{g}$ by $\mathrm{ad}\, X(Y) = [X, Y]$. Similarly, every $Z \in \mathbf{g}_c$ defines a complex–linear map (denoted by $\mathrm{ad}\, Z$) on $\mathbf{g}_c$. It therefore makes sense to look for eigenvalues and eigenvectors of $\mathrm{ad}\, Z$.
2. $\mathbf{g}$ has a (Lie) subalgebra $\mathbf{h}$ called a *Cartan subalgebra,* defined as a maximal abelian subalgebra of $\mathbf{g}$ satisfying an additional technical condition. $\mathbf{h}$ is analogous to a "maximal commuting set of observables" in quantum mechanics. Its complexification, denoted by $\mathbf{h}_c$, is a maximal abelian subalgebra of $\mathbf{g}_c$. (The technical condition on $\mathbf{h}$ essentially ensures that each of the linear maps $\mathrm{ad}\, H$, with $H \in \mathbf{h}_c$, has a complete set of eigenvectors in $\mathbf{g}_c$, hence that all the linear transformations $\mathrm{ad}\, H$ with $H \in \mathbf{h}_c$ can be diagonalized simultaneously.)
3. For every complex–linear form α: $\mathbf{h}_c \to \mathbb{C}$, let

$$\mathbf{g}^\alpha = \{Z \in \mathbf{g}_c \mid\ [H, Z] = \alpha(H)Z \quad \forall\, H \in \mathbf{h}_c\}. \tag{16}$$

   That is, $\mathbf{g}^\alpha$ is the set of all vectors in $\mathbf{g}_c$ which are common eigenvectors of $\mathrm{ad}\, H$ for every H in $\mathbf{h}_c$, with corresponding eigenvalue $\alpha(H)$. For generic α, no such eigenvectors will exist, hence $\mathbf{g}^\alpha = \{0\}$. When $\mathbf{g}^\alpha \neq \{0\}$, α is called a *root* (of $\mathbf{g}_c$ with respect to $\mathbf{h}_c$), each Z in $\mathbf{g}_c$ is called a *root vector* and $\mathbf{g}^\alpha$ is called a *root subspace* of $\mathbf{g}_c$. Clearly, the zero–form $\alpha(H) \equiv 0$ is a root, since the fact that $\mathbf{h}_c$ is abelian means that $\mathbf{g}^0$ contains $\mathbf{h}_c$. The fact that $\mathbf{h}_c$ is maximal–abelian means that actually $\mathbf{g}^0 = \mathbf{h}_c$, since every Z in $\mathbf{g}^0$ commutes with all H in $\mathbf{h}_c$.
4. Let $Z_\alpha \in \mathbf{g}^\alpha$ and $Z_\beta \in \mathbf{g}^\beta$, where α and β are arbitrary linear forms on $\mathbf{h}_c$. Then the Jacobi identity implies that for every H in $\mathbf{h}_c$,

$$\begin{aligned}[H, [Z_\alpha, Z_\beta]] &= [[H, Z_\alpha], Z_\beta] + [Z_\alpha, [H, Z_\beta]] \\ &= (\alpha(H) + \beta(H))\ [Z_\alpha, Z_\beta],\end{aligned} \tag{17}$$

   thus $[Z_\alpha, Z_\beta] \in \mathbf{g}^{\alpha+\beta}$. This statement is abbreviated as

$$\left[\mathbf{g}^\alpha, \mathbf{g}^\beta\right] \subset \mathbf{g}^{\alpha+\beta} \qquad \forall \alpha, \beta \in \mathbf{h}_c^*. \tag{18}$$

5. The set of all *nonzero* roots is denoted by Δ. $\mathbf{g}_c$ has a direct–sum decomposition *(root space decomposition)* as

$$\mathbf{g}_c = \mathbf{h}_c + \sum_{\alpha \in \Delta} \mathbf{g}^\alpha, \tag{19}$$

and each $\mathbf{g}^\alpha$ (with $\alpha \in \Delta$) is a one–dimensional subspace of $\mathbf{g}_c$.

6. The *Killing form* of $\mathbf{g}_c$ is the bilinear symmetric form defined by

$$B(Z, Z') = \text{trace}\,(\text{ad}\, Z\, \text{ad}\, Z')\,. \tag{20}$$

It is non–degenerate on $\mathbf{g}_c$, and its restriction to $\mathbf{h}_c$ is also non–degenerate. Since a non–degenerate bilinear form on a vector space defines an isomorphism between that space and its dual, the restriction of B to $\mathbf{h}_c$ defines an isomorphism between $\mathbf{h}_c$ and $\mathbf{h}_c^*$. The vector in $\mathbf{h}_c$ corresponding to $\alpha \in \mathbf{h}_c^*$ is denoted by H_α. It is defined by

$$B(H_\alpha, H) = \alpha(H) \qquad \forall H \in \mathbf{h}_c. \tag{21}$$

Note that the vector corresponding to $\alpha \equiv 0$ is $H_0 = 0$.

7. If $\alpha \in \Delta$, then $-\alpha \in \Delta$ and $\alpha(H_\alpha) \equiv B(H_\alpha, H_\alpha) \neq 0$. Furthermore,

$$\left[\mathbf{g}^\alpha, \mathbf{g}^{-\alpha}\right] = \mathbb{C}\, H_\alpha, \tag{22}$$

i.e. the set of all brackets $[Z_\alpha, Z_{-\alpha}]$ with $Z_\alpha \in \mathbf{g}^\alpha$ fills out the one–dimensional subspace spanned by H_α. ($H_\alpha \neq 0$ since $\alpha \in \Delta$.)

8. Any two root subspaces $\mathbf{g}^\alpha$ and $\mathbf{g}^\beta$ with $\alpha + \beta \neq 0$ are "orthogonal" with respect to B.
9. It is possible to choose (non–uniquely) a subset Δ^+ of Δ, called a set of *positive roots*, such that
   (a) If α and β belong to Δ^+ and $\alpha + \beta$ is a root, then $\alpha + \beta$ belongs to Δ^+.
   (b) The set $\Delta^- \equiv -\Delta^+$ is disjoint from Δ^+.
   (c) $\Delta = \Delta^+ \cup \Delta^-$.

   It follows from (4) and (9) that

$$\mathbf{n}^+ \equiv \sum_{\alpha \in \Delta^+} \mathbf{g}^\alpha \qquad \text{and} \qquad \mathbf{n}^- \equiv \sum_{\alpha \in \Delta^-} \mathbf{g}^\alpha \tag{23}$$

are Lie subalgebras of $\mathbf{g}_c$, and by (5),

$$\mathbf{g}_c = \mathbf{h}_c + \mathbf{n}^+ + \mathbf{n}^-. \tag{24}$$

Since there can only be a finite number of roots, (4) and (9) also imply that after taking a finite number of brackets of elements in either $\mathbf{n}^+$ or $\mathbf{n}^-$ we obtain zero; that is, the subalgebras $\mathbf{n}^\pm$ are *nilpotent*. We may choose a basis for $\mathbf{g}_c$ as follows: Pick a non–zero vector Z_α from

each $\mathbf{g}^\alpha$ with $\alpha \in \Delta^+$. Then by (7) there is a unique vector $Z_{-\alpha}$ in $\mathbf{g}^{-\alpha}$ with

$$[Z_\alpha, Z_{-\alpha}] = 2H_\alpha/\alpha(H_\alpha). \tag{25}$$

Since Z_α and $Z_{-\alpha}$ are root vectors, they also satisfy

$$\begin{aligned} [H_\alpha, Z_\alpha] &= \alpha(H_\alpha) Z_\alpha \\ [H_\alpha, Z_{-\alpha}] &= -\alpha(H_\alpha) Z_{-\alpha}, \end{aligned} \tag{26}$$

and $\alpha(H_\alpha) \neq 0$ (property (7)). If we define

$$C_\alpha = H_\alpha/\alpha(H_\alpha), \tag{27}$$

then each triplet $(Z_\alpha, C_\alpha, Z_{-\alpha})$ spans a Lie subalgebra of $\mathbf{g}_c$ which is isomorphic to $\mathbf{sl}(2, \mathbb{C})$:

$$[C_\alpha, Z_{\pm\alpha}] = \pm Z_{\pm\alpha}, \qquad [Z_\alpha, Z_{-\alpha}] = 2C_\alpha. \tag{28}$$

The root space decomposition then shows that $\mathbf{g}_c$ is a direct sum of copies of $\mathbf{sl}(2, \mathbb{C})$, indexed by Δ^+.

The connection with the non–hermitian combinations $X \pm iP$ can now be explained. The following argument is heuristic and has no pretense to rigor. Its sole purpose is to motivate the construction of general holomorphic frames. It will be made precise in section 3.6 at the level of unitary representations, where a clear geometric interpretation will be given.

Begin with the group $G = SU(2)$, which is a simple, real Lie group whose Lie algebra $\mathbf{g} = \mathbf{su}(2)$ has a basis $\{-iJ_1, -iJ_2, -iJ_3\}$ satisfying

$$\begin{aligned} [J_1, J_2] &= iJ_3 \\ [J_2, J_3] &= iJ_1 \\ [J_3, J_1] &= iJ_2. \end{aligned} \tag{29}$$

We first show that $\mathbf{w}$ can be obtained as a "contraction limit" of $\mathbf{g}$.* Choose a positive number κ and define $K_1 = \kappa J_1$, $K_2 = \kappa J_2$ and $K_3 = \kappa^2 J_3$. These form a new basis for $\mathbf{g}$, with

* The idea of group contractions is due to Inönü and Wigner [1953].

$$\begin{aligned}
[K_1, K_2] &= iK_3 \\
[K_2, K_3] &= i\kappa^2 K_1 \\
[K_3, K_1] &= i\kappa^2 K_2.
\end{aligned} \tag{30}$$

In the limit $\kappa \to 0$, $\mathbf{g}$ becomes isomorphic to $\mathbf{w}$. We will see in section 3.6 that within a unitary irreducible representation of $SU(2)$, the operator K_3 can be chosen so that $K_3 \to I$ as $\kappa \to 0$, hence we may interpret the limits of K_1 and K_2 as X and P, respectively.

Now apply the above structure theory to $\mathbf{g}_c$, which is just $\mathbf{sl}(2, \mathbb{C})$. The direct sum of copies of $\mathbf{sl}(2, \mathbb{C})$ therefore reduces to a single term. For the Cartan subalgebra of $\mathbf{g}$ we can choose $\mathbf{h} = \mathbb{R} J_3$ (the one–dimensional subspace spanned by J_3), so that $\mathbf{h}_c = \mathbb{C} J_3$. Two linearly independent root vectors are given by $J_\pm = J_1 \pm iJ_2$, with

$$[J_3, J_+] = J_+, \qquad [J_3, J_-] = -J_-. \tag{31}$$

We choose $\alpha(J_3) = 1$ as the single "positive" root and $J_\pm$ for the basis vectors of the two one–dimensional root subspaces. Then

$$[J_+, J_-] = 2J_3 \tag{32}$$

shows that C_α corresponds to J_3. Setting $K_\pm = \kappa J_\pm$ and $K_3 = \kappa^2 J_3$, and taking the limit $\kappa \to 0$, we obtain the correspondence

$$\begin{aligned}
K_+ &\to S + iT \mapsto X + iP = A \\
K_- &\to S - iT \mapsto X - iP = A^* \\
K_3 &\to E \mapsto I,
\end{aligned} \tag{33}$$

where we have used "$\mapsto$" to denote the representation of $\mathbf{w}$ on $L^2(\mathbb{R})$. [This correspondence is not unique; for example, $K_+ \to A^*$, $K_- \to A$, $K_3 \to -E$ is equally good. As will be shown in section 3.6, *both* of these correspondences actually occur as weak limits, due to the fact that an irreducible representation of $SU(2)$ contracts to a *reducible* representation of $\mathbf{w}$.] Note that the three roots $\{\kappa^2, 0, -\kappa^2\}$ all merge into a single root, zero, in the contraction limit. Thus the operators A and A^* are interpreted as the contraction limits of root vectors.

Note: Another way of seeing the importance and naturality of A and A^* is in the context of the Kirillov–Kostant–Souriau theory, sometimes called "Geometric Quantization," applied to the *Oscillator*

group (Streater [1967]; see section 3.6). For background on Geometric Quantization, see Kirillov [1976], Kostant [1970] and Souriau [1970]; see also Guillemin and Sternberg [1984], Simms and Woodhouse [1976] and Śniatycki [1980].

We are, at last, ready to generalize the construction of group–representations on spaces of holomorhic functions to other groups. We begin with a semisimple, real Lie group G and a unitary irreducible representation U of G on a Hilbert space $\mathcal{H}$. To avoid technical difficulties, we will here assume that G is *compact*. (The reason for assuming compactness, as well as ways to get around it, will be discussed below.) Then the irreducibility of U implies that $\mathcal{H}$ be finite–dimensional, hence the operators $U(g)$ representing group operations are just unitary matrices and the operators $U(X)$ representing elements of $\mathbf{g}$ are skew–adoint matrices. (Sometimes the operator representing X is written as $dU(X)$ to emphasize its "infinitesimal" nature; we will write it as $U(X)$ to keep the notation simple.) At the Lie algebra level, U extends, by complex–linearity, to a representation T of $\mathbf{g}_c$:

$$T(X + iY) \equiv U(X) + iU(Y). \tag{34}$$

T is the unique representation of $\mathbf{g}_c$ that extends U; that is, for all $c_1, c_2 \in \mathbb{C}$ and all $Z_1, Z_2 \in \mathbf{g}_c$,

$$\begin{aligned} T(c_1 Z_1 + c_2 Z_2) &= c_1 T(Z_1) + c_2 T(Z_2) \\ T([Z_1, Z_2]) &= [T(Z_1), T(Z_2)] \\ T(X) &= U(X) \qquad \text{if} \qquad X \in \mathbf{g} \subset \mathbf{g}_c. \end{aligned} \tag{35}$$

The matrices $T(Z)$ are no longer skew–adjoint, but because of the finite dimensionality of $\mathcal{H}$, there is no problem in exponentiating them to give a representation of G_c on $\mathcal{H}$, which we also denote by T. Thus, the representative of the group element $\exp Z$ of G_c is defined by*

$$T(\exp Z) = \exp[T(Z)], \qquad Z \in \mathbf{g}_c. \tag{36}$$

Note: If G is non–compact, $\exp Z$ is still well–defined but the right–hand side of the above equation is problematic since for any non–trivial representation U, $\mathcal{H}$ is infinite–dimensional and $T(Z)$ will, in

* Since G is compact, it is *exponential;* that is, every group element can be written in the form $\exp Z$ for some Z.

general, be a non–skew–adjoint, unbounded operator. (Even the definition $T(Z) = T(X) + iT(Y)$ becomes troublesome, since the skew–adjoint operators $T(X)$ and $T(Y)$ may both be unbounded and their domains may have little in common; additional assumptions must be made.) In the case of $\mathcal{W}$, this was resolved by restricting $\exp[T(Z)]$ to act on analytic vectors. A similar approach is used in extending the present construction to non–compact G. For the present, we continue to assume that G is compact to avoid this problem. #

Lemma 3.2..

(a) $g \mapsto T(g)$ is an irreducible (non–unitary) *representation of G_c.*

(b) The map $Z \mapsto T(\exp Z)$ is analytic as a map from the complex vector space $\mathbf{g}_c$ to the complex matrices on $\mathcal{H}$.

Proof. If A is a matrix which commutes with all $T(Z)$ for $Z \in \mathbf{g}_c$, then in particular it commutes with all $U(X)$ for $X \in \mathbf{g}$, hence must be a multiple of the identity since U is irreducible. Therefore T is irreducible. To prove (b), note that $Z \mapsto T(\exp Z)$ is, by definition, the composite of the two analytic maps $Z \mapsto T(Z)$ and $T(Z) \mapsto \exp[T(Z)]$. ∎

It follows that the map T from G_c (considered as a complex manifold; see Wells [1980]) to the group $GL(\mathcal{H})$ of non–singular matrices on $\mathcal{H}$ is also analytic; that is, T is a *holomorphic representation* of G_c, obtained by analytically continuing the representation U of G. Now the point of Onofri's construction is this: We have seen that by choosing a state which is stable under H, U can be reformulated as a representation of G on a space of functions $\tilde{f}(m)$ defined over the homogeneous space G/H. In the case $G = \mathcal{W}$, G/H was identified as a *phase space*, but in general it is not clear that it can be interpreted as such. Following Onofri, we will show that:

(a) The representation T induces a complex structure on the homogeneous space G/H, making it into a *complex manifold* on which G acts by *holomorphic transformations*. (Such a manifold is called a complex homogeneous space of G, or a *holomorphic* homogeneous G–space.)

(b) In addition, G/H has the (symplectic) structure of a classical phase space, and the action of G on G/H is by canonical transformations. Thus it becomes possible to think of G/H as *phase space*. To actually identify G/H as *the* phase space of a classical physical system, i.e. as the set of dynamical trajectories followed by that system, it is necessary for G to include the *dynamics* for the system, i.e. its evolution group, of which nothing has been said so far. This will be discussed in the next chapter.

The representatives $U(H)$ of the elements H of $\mathbf{h}$ form a commuting set of skew–adjoint matrices, hence can all be diagonalized simultaneously. Let h be a common eigenvector:

$$U(H)\,h = \lambda(H)\,h, \tag{37}$$

where $\lambda(H)$ is *imaginary*. Since U is linear at the Lie algebra level, it follows that λ is a linear functional on $\mathbf{h}$, called the *weight* of h. (Roots are simply weights in the adjoint representation, where $\mathcal{H}$ is replaced by $\mathbf{g}_c$ and $U(H)$ by $\mathrm{ad}H$.) For any non–zero element Z_α in $\mathbf{g}^\alpha$ with α in Δ^+, we have (remembering that $U(H) = T(H)$ since $H \in \mathbf{h}$)

$$\begin{aligned} T(H)T(Z_\alpha)\,h &= [T(H), T(Z_\alpha)]\ \ h + T(Z_\alpha)T(H)\,h \\ &= T\left([H, Z_\alpha]\right)\ \ h + \lambda(H)\,T(Z_\alpha)\,h \\ &= (\alpha(H) + \lambda(H))\ T(Z_\alpha)\ \ h. \end{aligned} \tag{38}$$

That is, $T(Z_\alpha)$ "raises" the weight by α. Similarly, for $\alpha \in \Delta^-$, $T(Z_\alpha)$ *lowers* the weight by $-\alpha$. Since non–zero vectors with different weights are linearly independent and $\mathcal{H}$ is finite–dimensional, it follows that $\mathcal{H}$ must contain a non–zero vector with *lowest weight*, i.e. such that

$$T(Z_\alpha)\,h = 0 \qquad \forall \alpha \in \Delta^-. \tag{39}$$

Equivalently,

$$T(Z)\,h = 0 \qquad \forall Z \in \mathbf{n}^-. \tag{40}$$

For the group $\mathcal{W}$, h was the "ground state" χ_0 and the above equations correspond to $T(-iE)\,\chi_0 = -i\chi_0$ (so $\lambda(-iE) = -i$) and $A\chi_0 = 0$.

Consider the subalgebra

$$\mathbf{b} = \mathbf{h}_c + \mathbf{n}^+ \tag{41}$$

of $\mathbf{g}_c$, called a *Borel subalgebra*. If $N \in \mathbf{n}^+$ and we denote by $\bar{N}$ its complex–conjugate with respect to the real subalgebra $\mathbf{g}$, then $\bar{N} \in \mathbf{n}^-$. Hence for arbitrary $Z = H + N \in \mathbf{b}$, we have

$$T(\bar{Z})h = T(\bar{H})h + T(\bar{N})h = T(\bar{H})h, \tag{42}$$

and $\bar{H}$ belongs to $\mathbf{h}_c$ since $[\bar{H}, \mathbf{h}_c] = 0$. Extending λ by complex–linearity to $\mathbf{h}_c$, we therefore have

$$T(\bar{Z})h = \lambda(\bar{H})\, h, \tag{43}$$

hence

$$\exp\left[T(\bar{Z})\right]\ h = \exp\left[\lambda(\bar{H})\right]\ h. \tag{44}$$

The subgroup $B = \exp(\mathbf{b})$ is called a *Borel subgroup* of $\mathbf{g}_c$. For Z as above, let $b = \exp Z \in B$, $\bar{b} = \exp \bar{Z}$ and $\pi(\bar{b}) = \exp[\lambda(\bar{H})]$. Then

$$T(\bar{b})\, h = \pi(\bar{b})\, h. \tag{45}$$

Since $\lambda(H)$ is imaginary for $H \in \mathbf{h}$, complex–linearity implies that $\lambda(\bar{H}) = -\overline{\lambda(H)}$ for $H \in \mathbf{h}_c$. Hence

$$\pi(\bar{b}) = \overline{\pi(b)}^{\,-1}, \qquad b \in B. \tag{46}$$

The map $\pi\colon B \to \mathbb{C}^*$ satisfies $\pi(b_1 b_2) = \pi(b_1)\,\pi(b_2)$, i.e. it is a character of B. Furthermore, since $\pi(b)$ is analytic in the group parameters of b, Onofri calls it a *holomorphic character* of B.

Notice that the *state* corresponding to h (i.e., the one–dimensional subspace spanned by it) is invariant under B. If we restrict ourselves to the *real* group G, this means that the state is invariant under the subgroup H. If H is the *maximal* subgroup of G leaving this state invariant, then the weight λ is called *non–singular*. In that case, H plays the same role as it did in the last section: it is the *stability subgroup* of the state. We will assume this to be the case; if it is not (in which case λ is *singular*), the present considerations still apply but in modified form. Note that in the non–singular case, the stability subgroup is *abelian*.

We now adapt the construction of the last section in a way which respects the complex–analytic structure of G_c. Introduce the notation

$$(T(g)^*)^{-1} \equiv T^{\#}(g), \qquad g \in G_c. \tag{47}$$

Since the representation U of G is unitary, $U(X)^* = -U(X)$ for $X \in \mathbf{g}$; hence for $Z \in \mathbf{g}_c$, we have $T(Z)^* = -T(\bar{Z})$. It follows that for group elements $g = \exp(Z)$ of G_c,

$$T(g)^{\#} = T(\bar{g}), \qquad g \in G_c, \tag{48}$$

where $\bar{g} = \exp(\bar{Z})$ and, in particular, $T(g)^{\#} = T(g) = U(g)$ for $g \in G$. Define the vectors

$$h_g = T(g)^{\#}\, h, \qquad g \in G_c, \tag{49}$$

which, when restricted to $g \in G$, coincide with the earlier frame vectors but are anti–holomorphic in the group parameters of G_c. An arbitrary vector $f \in \mathcal{H}$ defines a holomorphic function on G_c by

$$\tilde{f}(g) \equiv \langle\, h_g \,|\, f \,\rangle. \tag{50}$$

For arbitrary $b \in B$,

$$h_{gb} = T(g)^{\#} T(b)^{\#} h = \overline{\pi(b)}^{\,-1} h_g, \tag{51}$$

hence

$$\tilde{f}(gb) = \pi(b)^{-1} \tilde{f}(g). \tag{52}$$

Note: The reader familiar with fiber bundles (Kobayashi and Nomizu [1963, 1969]) will recognize the above equation as the condition defining a holomorphic section of the holomorphic line bundle associated to the principal bundle $B \to G_c \to G_c/B$ by the character $\pi\colon B \to \mathbb{C}^*$. We now proceed to construct this section in a naive way, that is, without assuming any knowledge of bundle theory. #

The above shows that the *state* determined by h_g depends only on the left coset gB, which we denote by z. Let

$$\mathcal{Z} = G_c/B \tag{53}$$

be the left coset space. Now
(a) G_c is a complex manifold;
(b) B, as a complex subgroup, is a complex submanifold of G_c;
(c) the projection map $G_c \to \mathcal{Z}$ is holomorphic.
Hence it follows that $\mathcal{Z}$ is a complex manifold. This means that a neighborhood of each point $z = gB \in \mathcal{Z}$ can be parametrized by a local chart, i.e. a set of local complex coordinates $(z_1, \ldots, z_n)$ (say, with $(0, \ldots, 0)$ corresponding to z), and the transformation from one local chart to another on overlapping neighborhoods is a local holomorphic function. In the case of $G = \mathcal{W}$, B corresponds (under the contraction limit) to the complex subalgebra spanned by E and A, and $\mathcal{Z}$ can be identified with $\mathbb{C}$, hence only a single chart is needed to cover all of $\mathcal{Z}$. In general, more than one chart is necessary. For $G = SU(2)$, we will see that $\mathcal{Z}$ is the Riemann sphere S^2, hence two charts are needed; however, the north pole has measure zero, and a single chart will do for $S^2 \backslash \{\infty\} \approx \mathbb{C}$.

Since G_c acts on itself by holomorphic transformations, its action on $\mathcal{Z}$ is also by holomorphic transformations. This means the following: For $g_1 \in G_c$ and $z = gB \in \mathcal{Z}$, let $w(z) = g_1 z \equiv (g_1 g)B$. If $(z_1, \ldots, z_n)$ and $(w_1, \ldots, w_n)$ are local charts in neighborhoods of z and w, respectively, then the mapping ϕ: $(z_1, \ldots, z_n) \mapsto (w_1, \ldots, w_n)$ is holomorphic in a neighborhood of $(0, \ldots, 0)$. (It must, of course, be locally invertible with holomorphic inverse.)

We could proceed as in section 3 and consider the states

$$P_z = \frac{|\, h_g \,\rangle\langle\, h_g \,|}{\langle\, h_g \,|\, h_g \,\rangle}, \qquad z \equiv gB \in \mathcal{Z}, \tag{54}$$

where we must now divide by $\|h_g\|^2$ since the non–unitary operator $T(g)$ does not preserve norms. However, this would spoil the holomorphy which we are attempting to study. Instead, proceed as follows: Choose an arbitrary reference point a in G_c and define the *holomorphic coherent states*

$$\chi_z^a = \frac{h_g}{\langle\, h_a \,|\, h_g \,\rangle}. \tag{55}$$

As indicated by the notation, the right–hand side depends only on the coset $z = gB$. There is no guarantee that the denominator on the right–hand side is non–zero, but certainly the open set

$$U_\alpha = \{g \in G_c \mid \langle\, h_a \,|\, h_g \,\rangle \neq 0\} \tag{56}$$

(which, as indicated, depends only on the coset $\alpha = aB$) contains a, and its projection V_α to $\mathcal{Z}$ is an open set containing α such that χ_z^a is defined for all z in V_α. Hence by choosing more than one reference point a, if necessary, we can cover $\mathcal{Z}$ with patches V_α on which the χ_z^a's are defined.

An arbitrary vector $f \in \mathcal{H}$ can now be expressed as a *local holomorphic function*

$$\tilde{f}^a(z) = \langle\, \chi_z^a \,|\, f \,\rangle \tag{57}$$

of z in V_α, with *transition functions*

$$\begin{aligned} \tilde{f}^b(z) &= \frac{\langle\, h_g \,|\, h_a \,\rangle}{\langle\, h_g \,|\, h_b \,\rangle}\, \tilde{f}^a(z) \\ &\equiv \tau_a^b(z) \tilde{f}^a(z), \qquad z \in V_\alpha \chi^{\alpha+1} V_\beta. \end{aligned} \tag{58}$$

The reader may wonder where this is all leading, since we are ultimately interested in the *real* group G and not in G_c. Here is the point: It is known (Bott [1957]) that the complex homogeneous space $\mathcal{Z} = G_c/B$ actually coincides with the *real* homogeneous space $M = G/H$ used in Perelomov's construction! For example, consider the Weyl–Heisenberg group: G/H is parametrized by (x, p) while G_c/B is parametrized by $x - ip$, and they are the same set but with the difference that the latter has gained a complex structure. The identification of M with $\mathcal{Z}$ in general can be obtained by noting that *as a subgroup of G_c, G acts on $\mathcal{Z}$ by holomorphic transformations*; this action turns out to be transitive, and the isotropy subgroup at the "origin" $z_0 = B$ is H, hence $\mathcal{Z} \approx G/H$. In other words, *M inherits a complex structure from G_c, and the natural action of G on M preserves this structure.* Because of this, we need not deal directly with G_c to reap the benefits of the complex structure. Let us therefore restrict ourselves to G. Then

$$h_g = \langle\, h_a \,|\, h_g \,\rangle\, \chi_z^a = U(g)\, h, \tag{59}$$

hence $\| h_g \| = \| h \| \equiv 1$ and the state corresponding to h_g is

$$\begin{aligned} |\, h_g \,\rangle\langle\, h_g \,| \; &= |\langle\, h_a \,|\, h_g \,\rangle|^2 \; |\, \chi_z^a \,\rangle\langle\, \chi_z^a \,| \\ &\equiv e^{-\phi(z,\alpha)} \, |\, \chi_z^a \,\rangle\langle\, \chi_z^a \,|\,, \end{aligned} \tag{60}$$

where $\phi(z, \alpha) \equiv -2 \ln |\langle\, h_a \,|\, h_g \,\rangle|$ depends only on $z = gH$ and $\alpha = aH$ and their complex conjugates (it is *not* analytic). Notice that in eq. (60), *the left–hand side, hence also the right–hand side, is independent of a.* Only the three individual factors on the right–hand side depend on a. This becomes important if several patches are needed to cover $\mathcal{Z}$, since it means that we can change the reference point without affecting the smoothness of the frame. With the above definitions, the resolution of unity derived in section 3 becomes

$$\int_M d\mu_M(z) \; e^{-\phi(z,\alpha)} \, |\, \chi_z^a \,\rangle\langle\, \chi_z^a \,| \; = I, \tag{61}$$

where we have assumed for simplicity that a single chart suffices. If more than one chart is needed, partition M as a (disjoint) union $\cup_n M_n$, where each M_n is covered by a single chart. Since, by the above remark, the integrand is independent of a, the corresponding integrals I_n form a *partition of unity* in the sense that $\sum_n I_n = I$. Therefore, the holomorphic coherent states form a tight frame which we call the

holomorphic G–frame associated with the representation U and the lowest–weight vector χ. In this connection, note that a choice of lowest weight actually *determines* the representation U up to equivalence, hence a more economical terminology would be to call the above the holomorphic G–frame associated with χ. The corresponding inner product is given by

$$\langle f_1 \,|\, f_2 \rangle = \int_M d\mu_M(z)\, e^{-\phi(z,\alpha)}\, \overline{\tilde{f}_1^a(z)}\, \tilde{f}_2^a(z). \tag{62}$$

In the case $G = \mathcal{W}$, since $\mathcal{Z} = \mathbb{C}$, everything can be done globally: Just one chart is needed, and the reference point a can be fixed once for all. Taking $a = 1$ (the identity element of G) and $\alpha = 0$, we find that the χ_z's reduce to the canonical coherent states and

$$e^{-\phi(z,0)} = |\langle \chi_z \,|\, \chi_0 \rangle|^2 = e^{-\bar{z}z/2}. \tag{63}$$

Hence in the general case, $e^{-\phi}$ takes the place of the Gaussian weight function: it corrects for the fact that *holomorphic translations do not preserve the norm.* The action of G on the space of local holomorphic functions $\tilde{f}^a(z)$ has a "multiplier":

$$\begin{aligned} \langle \chi_z^a \,|\, U(g_1) f \rangle &= \frac{\langle h_g \,|\, U(g_1)\, f \rangle}{\langle h_g \,|\, h_a \rangle} \\ &= \frac{\langle h_{g_1^{-1}g} \,|\, h_a \rangle}{\langle h_g \,|\, h_a \rangle} \cdot \frac{\langle h_{g_1^{-1}g} \,|\, f \rangle}{\langle h_{g_1^{-1}g} \,|\, h_a \rangle} \\ &\equiv \gamma(z, g_1, \bar{\alpha}) \tilde{f}^a({g_1}^{-1} z), \end{aligned} \tag{64}$$

where $z = gH$, $\alpha = aH$ and

$$\begin{aligned} \gamma(z, g_1, \bar{\alpha}) &= \frac{\langle h_g \,|\, U(g_1)\, h_a \rangle}{\langle h_g \,|\, h_a \rangle} \\ &= \frac{\langle h_{g_1^{-1}g} \,|\, h_a \rangle}{\langle h_g \,|\, h_a \rangle} \\ &= \frac{\langle h_g \,|\, h_{g_1 a} \rangle}{\langle h_g \,|\, h_a \rangle} \end{aligned} \tag{65}$$

is holomorphic in z and anti–holomorphic in α.

We have mentioned that M inherits a complex structure from G_c. Actually, this is only part of the story. Let ∂ and $\bar{\partial}$ denote the

external derivatives with respect to z and $\bar{z}$, respectively, i.e., in local coordinates,

$$\partial f = \frac{\partial f}{\partial z_k}\, dz_k, \quad \bar{\partial} f = \frac{\partial f}{\partial \bar{z}_k}\, d\bar{z}_k \tag{66}$$

(summation over k is implied). Consider the 2–form

$$\omega = i\partial\bar{\partial}\phi = i\frac{\partial^2 \phi}{\partial z_j \partial \bar{z}_k}\, dz^j \wedge d\bar{z}^k, \tag{67}$$

where ϕ is the function in the exponent of the weight function above.

Theorem.
(a) ω is closed, i.e. $\partial\omega = \bar{\partial}\omega = 0$,
(b) ω is independent of the reference point a,
(c) ω is invariant under the action of G, and
(d) ω is non–degenerate, if λ is non–singular.

Proof. We prove (a), (b) and (c). For the proof of (d), see Onofri [1975].
(a) follows from the fact that $\bar{\partial}+\partial = d$ is the total exterior derivative, hence the identity $d^2 = 0$ implies

$$\partial^2 = 0, \quad \bar{\partial}^2 = 0, \quad \bar{\partial}\partial + \partial\bar{\partial} = 0. \tag{68}$$

(c) follows from the fact that $\gamma(z, g_1, \bar{\alpha})$ is holomorphic in z, hence $|\gamma|^2$ is harmonic and $\partial\bar{\partial}|\gamma|^2 = 0$. But eq. (65) shows that

$$|\gamma(z, g_1, \alpha)|^2 = \frac{\exp[-\phi(g_1^{-1}z, \alpha)]}{\exp[-\phi(z, \alpha)]}, \tag{69}$$

which implies that the pullback $g_1^*\omega$ of ω under g_1 equals ω, i.e. that ω is invariant.
(b) follows from (c) and $\phi(z, g_1\alpha) = \phi(g_1^{-1}z, \alpha)$. ∎

The property (b) implies that ω is defined *globally* on $\mathcal{Z}$, whereas (a) and (d) mean that ω is a *symplectic form* (Kobayashi and Nomizu [1969]) on $\mathcal{Z}$, which makes $\mathcal{Z}$ a possible classical phase space. Finally, (c) means that the symplectic structure defined by ω is G–invariant, hence G acts on $\mathcal{Z}$ by *canonical transformations*. (Actually, the 2–form ω together with the complex structure define a *Kähler structure* on $\mathcal{Z}$, i.e. a Hermitian metric such that the complex structure is invariant under parallel translations.)

3.5. The Rotation Group

A simple but important example of the foregoing methods is provided by their application to the three–dimensional rotation group $SO(3)$. The resulting frame vectors are known as *spin coherent states*. An excellent and detailed account of this is given in Perelomov [1986]; the treatment here will be fairly brief. $SO(3)$ is locally isomorphic to the group $SU(2)$ of unitary unimodular 2×2 matrices, which we denote by G in this section. This is the set of all matrices

$$g = \begin{pmatrix} \alpha & \beta \\ -\bar{\beta} & \bar{\alpha} \end{pmatrix}, \qquad |\alpha|^2 + |\beta|^2 = 1, \tag{1}$$

hence $G \approx S^3$ (the unit sphere in $\mathbb{R}^4$) as a manifold. The Lie algebra $\mathbf{g}$ has a basis $\{J_1, J_2, J_3\}$ satisfying $[J_1, J_2] = iJ_3$ plus cyclic permutations (where, as usual, it is actually iJ_k which span the real algebra $\mathbf{g}$) which can be conveniently given as $J_k = (1/2)\sigma_k$ in terms of the Pauli matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{2}$$

Root vectors in $\mathbf{g}_c$ are given by $J_\pm = J_1 \pm iJ_2$, satisfying

$$[J_3, J_\pm] = \pm J_\pm, \qquad [J_+, J_-] = 2J_3. \tag{3}$$

The vectors $\{J_3, J_\pm\}$ form a complex basis for $\mathbf{g}_c$, with

$$J_+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad J_- = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}. \tag{4}$$

Unitary irreducible representations of G are characterized by a single number (highest weight) $s = 0, 1/2, 1, 3/2, \ldots$, with the representation space $\mathcal{H}$ having dimensionality $2s + 1$. The generators J_k are represented by hermitian matrices S_k satisfying the irreducibility (Casimir) condition

$$\mathbf{S}^2 \equiv S_1^2 + S_2^2 + S_3^2 = (1/2)(S_+S_- + S_-S_+) + S_3^2 = s(s+1). \tag{5}$$

A basis for $\mathcal{H}$ is obtained by starting with a highest–weight vector v_s, i.e.,

$$S_+ v_s = 0, \qquad S_3 v_s = s v_s, \tag{6}$$

and applying S_- repeatedly until the commutation relations imply that the resulting vector vanishes. This results (after normalization) in an orthonormal basis $\{v_s, v_{s-1}, \dots, v_{-s}\}$ satisfying

$$\begin{aligned} S_3 v_m &= m v_m \\ S_+ v_m &= \sqrt{(s-m)(s+m+1)}\, v_{m+1} \\ S_- v_m &= \sqrt{(s+m)(s-m+1)}\, v_{m-1}. \end{aligned} \tag{7}$$

To build a homogeneous frame as in section 3.3, we use a decomposition of G in terms of Euler angles,

$$g(\phi, \theta, \psi) = \exp(-i\phi J_3)\, \exp(-i\theta J_2)\, \exp(-i\psi J_3), \tag{8}$$

with $0 \le \phi < 2\pi$, $0 \le \theta \le \pi$ and $0 \le \psi < 4\pi$, which gives a corresponding decomposition of $U(g)$ as

$$U(\phi, \theta, \psi) = \exp(-i\phi S_3)\, \exp(-i\theta S_2)\, \exp(-i\psi S_3). \tag{9}$$

If $2s$ is odd, ϕ, θ and ψ have the same ranges as before; if $2s$ is even, then $\psi + 2\pi$ and ψ give the same operators, hence $0 \le \psi < 2\pi$. If we choose one of the basis vectors v_m as our initial vector h, then the stability subgroup is $H = \{g(0, 0, \psi)\} \approx S^1$. The homogeneous space $G/H \approx S^3/S^1$ is parametrized by (ϕ, θ), or by the unit vectors $\mathbf{n} = (\cos\phi \sin\theta, \sin\phi \sin\theta, \cos\theta)$, hence G/H can be identified with the unit sphere S^2. Choosing the section $\sigma\colon S^2 \to G$ as $\sigma(\phi, \theta) = (\phi, \theta, 0)$, we obtain the frame vectors

$$h_{\mathbf{n}} = e^{-i\phi S_3}\, e^{-i\theta S_2}\, h. \tag{10}$$

The G–invariant measure on S^2 is just the area measure

$$d\mathbf{n} = \sin\theta\, d\theta\, d\phi\,, \tag{11}$$

and the tight frame is

$$\int_{S^2} d\mathbf{n}\, |\, h_{\mathbf{n}}\, \rangle\langle\, h_{\mathbf{n}}\, | \;= cI \tag{12}$$

for some number c. To find c, take the trace of both sides and use the fact that $|\, h_{\mathbf{n}}\, \rangle\langle\, h_{\mathbf{n}}\, |$ is a rank–one projection operator (and hence its trace is unity). This gives

$$4\pi = c\ \mathrm{Tr}\, I = c(2s+1), \tag{13}$$

hence the resolution of unity is

$$\frac{2s+1}{4\pi}\int_{S^2} d\mathbf{n}\,|\,h_{\mathbf{n}}\,\rangle\langle\,h_{\mathbf{n}}\,|\,=I. \tag{14}$$

The overlap between frame vectors can be shown to be

$$|\,\langle\,h_{\mathbf{n}}\,|\,h_{\mathbf{n}'}\,\rangle\,|^{\,2} = \left(\frac{1+\mathbf{n}\cdot\mathbf{n}'}{2}\right)^{2s}, \tag{15}$$

hence they are orthogonal if and only if $\mathbf{n}' = -\mathbf{n}$.

To construct a holomorphic frame, we must consider the complex Lie algebra $\mathbf{g}_c$ and its Lie group $G_c = SL(2,\mathbb{C}\,)$ of unimodular 2×2 complex matrices

$$g = \begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix}, \qquad \alpha\delta - \beta\gamma = 1. \tag{16}$$

The Lie algebra $\mathbf{h}$ of the subgroup H of G used above is a Cartan subalgebra of $\mathbf{g}$, and $\mathbf{h}_c = \mathbb{C}J_3$. The corresponding root–space decomposition is

$$\mathbf{g}_c = \mathbf{h}_c + \mathbf{n}^+ + \mathbf{n}^- = \mathbb{C}J_3 + \mathbb{C}J_+ + \mathbb{C}J_-, \tag{17}$$

and this yields a *Gaussian* decomposition of (almost all) elements of G_c as products of lower–triangular, diagonal and upper–triangular matrices:

$$\begin{aligned} g(\zeta,d,\xi) &= e^{\zeta J_-}\, e^{2dJ_3}\, e^{\xi J_+}\\ &= \begin{pmatrix}1 & 0\\ \zeta & 1\end{pmatrix}\begin{pmatrix}e^d & 0\\ 0 & e^{-d}\end{pmatrix}\begin{pmatrix}1 & \xi\\ 0 & 1\end{pmatrix}\\ &\equiv \zeta_-\ \mathbf{d}\ \xi_+. \end{aligned} \tag{18}$$

If we write $N^{\pm} = \exp(\mathbf{n}^{\pm})$, then the above decomposition is $G_c \sim N^- H_c N^+$. Comparison with the original form of g gives

$$\begin{aligned} \alpha = e^d, \qquad & \beta = e^d\,\xi,\\ \gamma = \zeta e^d, \qquad & \delta = \zeta e^d\,\xi + e^{-d}. \end{aligned} \tag{19}$$

We call $\zeta(g)$, $\xi(g)$ and $e^d \equiv \alpha(g)$ the *Gaussian parameters* of g. Thus

$$\alpha(g) = \alpha, \quad \zeta(g) = \gamma/\alpha, \quad \xi(g) = \beta/\alpha. \tag{20}$$

Remarks:

1. Matrices with $\alpha = 0$, i.e. of the form

$$g = \begin{pmatrix} 0 & \beta \\ -\beta^{-1} & \delta \end{pmatrix}, \tag{21}$$

clearly do not have this decomposition. They form a 2–dimensional complex submanifold of G_c, hence have (group–) measure zero.

2. Elements of G are distinguished by $\delta = \bar{\alpha}$ and $\gamma = -\bar{\beta}$, which implies

$$\bar{\alpha}\alpha = (1 + \bar{\zeta}\zeta)^{-1} = (1 + \bar{\xi}\xi)^{-1}. \tag{22}$$

The Borel subgroup discussed in section 3.4 is $B = H_c N^+$ and consists of all matrices of the form

$$b = \begin{pmatrix} \alpha & \beta \\ 0 & \alpha^{-1} \end{pmatrix}. \tag{23}$$

Its cosets $\zeta_- B$ can therefore be parametrized by $\zeta \in \mathbb{C}$. The unattainable matrices with $\alpha = 0$ form a single coset, corresponding to $\zeta = \infty$. Hence the homogeneous space G_c/B is the Riemann sphere:

$$\mathcal{Z} \equiv G_c/B \approx \mathbb{C} \cup \{\infty\} \approx S^2, \tag{24}$$

in agreement with our earlier $G/H = S^2$ but with an added complex structure, as claimed in section 3.4. The action of G_c on $\mathcal{Z}$ is as follows: If $g\zeta'_- B = \zeta''_- B$, i.e.

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \zeta' & 1 \end{pmatrix} B = \begin{pmatrix} 1 & 0 \\ \zeta'' & 1 \end{pmatrix} B, \tag{25}$$

then

$$\zeta'' = \zeta(g\zeta'_-) = \frac{\gamma + \delta\zeta'}{\alpha + \beta\zeta'}. \tag{26}$$

That is, G_c acts on $\mathcal{Z}$ by *Möbius transformations*.

We defined $h_g = T(g)^{\#} h$, where $T(g)$ was the analytic continuation of $U(g)$ to G_c and $T(g)^{\#} = (T(g)^*)^{-1}$. But

$$\begin{aligned} T(\zeta_- \, \mathbf{d} \, \xi_+)^{\#} &= T(\zeta_+)^{\#} \, T(\mathbf{d})^{\#} \, T(\xi_-)^{\#} \\ &= \exp(-\bar{\zeta} S_+) \exp(-2\bar{d} S_3) \exp(-\bar{\xi} S_-). \end{aligned} \tag{27}$$

Hence the unique state left invariant by B is that corresponding to the vector of lowest weight, $h = v_{-s}$. For this choice, we get

$$h_g = e^{2\bar{d}s} \exp(-\bar{\zeta} S_+) \, h \equiv e^{2\bar{d}s} \, h_\zeta. \tag{28}$$

The holomorphic coherent states with respect to the reference point $g_1 \in G_c$ are then

$$\chi_\zeta^{g_1} = \frac{h_g}{\langle \, h_{g_1} \, | \, h_g \, \rangle} = \frac{e^{-2sd_1} h_\zeta}{\langle \, h_{\zeta_1} \, | \, h_\zeta \, \rangle}, \tag{29}$$

with

$$\langle \, h_{\zeta_1} \, | \, h_\zeta \, \rangle = \langle \, h \, | \, \exp(-\zeta_1 S_-) \, \exp(-\bar{\zeta} S_+) \, h \, \rangle. \tag{30}$$

To evaluate this, express $\exp(-\zeta_1 S_-) \, \exp(-\bar{\zeta} S_+)$ in the *opposite* factorization $N^+ H_c N^-$ and use $S_- h = 0$. It suffices to do the computation at the level of $2{\times}2$ matrices since the result depends only on the commutation relations, which are preserved by the representation. Thus

$$\begin{aligned} &\begin{pmatrix} 1 & 0 \\ -\zeta_1 & 1 \end{pmatrix} \begin{pmatrix} 1 & -\bar{\zeta} \\ 0 & 1 \end{pmatrix} \\ &\qquad = \begin{pmatrix} 1 & \xi' \\ 0 & 1 \end{pmatrix} \begin{pmatrix} e^{d'} & 0 \\ 0 & e^{-d'} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \zeta' & 1 \end{pmatrix} \equiv \xi'_+ \mathbf{d}' \zeta'_- \ , \end{aligned} \tag{31}$$

which gives

$$e^{-d'} = 1 + \zeta_1 \bar{\zeta}, \quad e^{-d'} \zeta' = -\zeta_1, \quad \xi' e^{-d'} = -\bar{\zeta}. \tag{32}$$

Hence

$$\begin{aligned} \langle \, h_{\zeta_1} \, | \, h_\zeta \, \rangle &= \langle \, h \, | \, T(\zeta'_+ \mathbf{d}' \zeta'_-) \, h \, \rangle \\ &= e^{-2sd'} = \left(1 + \zeta_1 \bar{\zeta}\right)^{2s}, \end{aligned} \tag{33}$$

and

$$\chi_\zeta^{g_1} = e^{-2sd_1} \left(1 + \zeta_1 \bar{\zeta}\right)^{-2s} h_\zeta. \tag{34}$$

Since a single chart covers all of $\mathcal{Z} \approx S^2$ except for the point at $\zeta = \infty$, just one reference point g_1 is needed in this case. The simplest choice is $g_1 = 1$, the identity of G, which gives

$$\chi_\zeta^1 = h_\zeta. \tag{35}$$

With this choice, the weight function is

$$\begin{aligned} e^{-\phi(\zeta)} &= |\langle h | h_g \rangle|^2, \quad g = \zeta_- \mathbf{d}\xi_+ \\ &= e^{2s(d+\bar{d})} |\langle h | h_\zeta \rangle|^2 \\ &= e^{2s(d+\bar{d})}. \end{aligned} \tag{36}$$

But for $g \in G$ we have the constraint

$$e^{2s(d+\bar{d})} = |\alpha|^2 = (1 + \bar{\zeta}\zeta)^{-1}, \tag{37}$$

hence

$$e^{-\phi(\zeta)} = (1 + \bar{\zeta}\zeta)^{-2s}. \tag{38}$$

To find the invariant measure on G_c/B, recall that the 2–form $\omega = i\partial\bar{\partial}\phi$ is invariant under G_c and non–degenerate. Hence it defines an invariant measure, once we choose a positive orientation on G_c/B. An easy computation gives

$$\begin{aligned} \omega &= -2is\partial\bar{\partial} \ln(1 + \bar{\zeta}\zeta) \\ &= \frac{2is d\bar{\zeta} \wedge d\zeta}{(1 + \bar{\zeta}\zeta)^2}. \end{aligned} \tag{39}$$

Thus

$$2si \int_{\mathbb{C}} \frac{d\bar{\zeta} \wedge d\zeta}{(1 + \bar{\zeta}\zeta)^{2s+2}} |h_\zeta\rangle\langle h_\zeta| = cI \tag{40}$$

for some c. Taking the trace and using

$$\langle h_\zeta | h_\zeta \rangle = (1 + \bar{\zeta}\zeta)^{2s} \tag{41}$$

gives, with $\zeta = re^{i\theta}$,

$$8\pi s \int_0^\infty \frac{r dr}{(1+r^2)^2} = c(2s+1), \tag{42}$$

from which $c = 4\pi s/(2s+1)$. Thus we have the resolution of unity

$$\frac{2s+1}{\pi} \int_{\mathbb{C}} \frac{d^2\zeta}{(1+\bar{\zeta}\zeta)^{2s+2}} \, |\, h_\zeta \,\rangle\langle\, h_\zeta \,| \, = I, \tag{43}$$

where $d^2\zeta$ is now Lebesgue measure on $\mathbb{C}$.

What do the functions $\tilde{f}(\zeta) = \langle\, h_\zeta \,|\, f \,\rangle$ look like? Consider the vectors

$$u_n = \frac{1}{n!}(-S_+)^n \, h, \quad n = 0, 1, 2, \cdots. \tag{44}$$

Then $u_0 \equiv h, u_1, \ldots, u_{2s}$ are linearly independent and $u_{2s+1} = 0$, since $S_+^{2s+1} = 0$. Thus

$$\begin{aligned} h_\zeta &= e^{-\bar{\zeta} S_+} h \\ &= u_0 + \bar{\zeta} u_1 + \cdots + (\bar{\zeta})^{2s} u_{2s}, \end{aligned} \tag{45}$$

hence

$$\tilde{f}(\zeta) = f_0 + \zeta f_1 + \cdots + \zeta^{2s} f_{2s}, \tag{46}$$

where $f_n = \langle\, u_n \,|\, f \,\rangle$. Thus, $\tilde{f}(\zeta)$ is a polynomial of degree $\leq 2s$ in ζ.

How are the two sets of frame vectors $h_{\mathbf{n}}$ and h_ζ related? It turns out that $\mathbf{n}$ is related to ζ by stereographic projection of the 2–sphere onto the complex plane. The exact relation depends on the particular factorizations of G used to construct the frames. In the case of $h_{\mathbf{n}}$, this was the Euler angle decomposition, whereas for h_ζ it was the Gaussian decomposition. However, there is an *intrinsic* way of relating the two sets of vectors, which goes as follows: Consider the functions

$$\tilde{S}_k(\zeta) = \frac{\langle\, h_\zeta \,|\, S_k h_\zeta \,\rangle}{\langle\, h_\zeta \,|\, h_\zeta \,\rangle}, \tag{47}$$

which are the *expectations* of the observables S_k in the state of h_ζ, and which correspond to the components of the classical angular momentum of a system whose only degrees of freedom are the rotational

motions given by G. $\tilde{S}_k(\zeta)$ is a quantum–mechanical version of a statistical average. Since $h_\zeta = \exp(-\bar{\zeta}S_+)h$, we have

$$\begin{aligned} \tilde{S}_+(\zeta) &= -\frac{\partial}{\partial\bar{\zeta}} \ln\langle\, h_\zeta \,|\, h_\zeta \,\rangle \\ &= -\frac{2s\zeta}{1+\bar{\zeta}\zeta}. \end{aligned} \tag{48}$$

To find $\tilde{S}_3(\zeta)$, note that $[S_3, S_+] = S_+$ implies $[S_3, S_+^n] = nS_+^n$, hence

$$\left[S_3, e^{-\bar{\zeta}S_+}\right] = -\bar{\zeta}S_+ e^{-\bar{\zeta}S_+} \tag{49}$$

and

$$\begin{aligned} S_3 h_\zeta &= -\bar{\zeta}S_+ h_\zeta + e^{-\bar{\zeta}S_+} S_3 h \\ &= -\left(\bar{\zeta}S_+ + s\right) h_\zeta. \end{aligned} \tag{50}$$

Hence

$$\tilde{S}_3(\zeta) = -s - \bar{\zeta}\tilde{S}_+(\zeta) = s\left[\frac{\bar{\zeta}\zeta - 1}{\bar{\zeta}\zeta + 1}\right]. \tag{51}$$

The equator of S^2 corresponds to $S_3 = 0$, hence to $|\zeta| = 1$, and the south pole ($S_3 = -s$) and north pole ($S_3 = s$) to $\zeta = 0$ and $\zeta = \infty$, respectively. The above equations imply that

$$\begin{aligned} \tilde{\mathbf{S}}^2(\zeta) &\equiv \tilde{S}_1(\zeta)^2 + \tilde{S}_2(\zeta)^2 + \tilde{S}_3(\zeta)^2 \\ &= |\tilde{S}_+(\zeta)|^2 + \tilde{S}_3(\zeta)^2 \\ &= s^2, \end{aligned} \tag{52}$$

thus $\tilde{\mathbf{S}}(\zeta)$ belongs to the 2–sphere of radius s centered at the origin. In fact, the correspondence $\zeta \leftrightarrow \tilde{\mathbf{S}}(\zeta)$ is a bijection if we include $\zeta = \infty$. The relation

$$\frac{\tilde{S}_+(\zeta)}{s - \tilde{S}_3(\zeta)} = -\zeta \tag{53}$$

shows that $-\zeta$ *is just the stereographic projection of* $\tilde{\mathbf{S}}(\zeta)$ *from the north pole to the complex plane tangent to the south pole.* We could choose $\mathbf{s}$ as a new independent variable ranging over the 2–sphere of

radius s minus the north pole and define $h_{\mathbf{s}} = h_\zeta$, where ζ is the unique point with $\tilde{\mathbf{S}}(\zeta) = \mathbf{s}$. Then the vectors $h_{\mathbf{s}}$ are essentially equivalent to the $h_{\mathbf{n}}$'s. Note that they are eigenvectors of the operator $\mathbf{s} \cdot \mathbf{S}$ with eigenvalue s^2, i.e.,

$$\mathbf{s} \cdot \mathbf{S} h_{\mathbf{s}} = s^2 h_{\mathbf{s}}, \tag{54}$$

since the equation obviously holds for $\mathbf{s} = (0, 0, -s)$, hence for all $\mathbf{s}$ by symmetry.

3.6. The Harmonic Oscillator as a Contraction Limit

In section 3.4, we gave a heuristic argument suggesting that the Weyl–Heisenberg group $\mathcal{W}$ is a "contraction limit" of $SU(2)$. This can now be made precise and given a geometric interpretation at the representation level. Furthermore, imitating the analysis of the non–relativistic limit of Klein–Gordon theory (next chapter), we shall gain an understanding of the relation between the Harmonic Oscillator and the canonical coherent states in the bargain. This connection between the rotation group and the harmonic oscillator has some potentially important applications in quantum field theory, which I hope to explore in the future.

We will study the limit of the representation of $G = SU(2)$ with spin s as $s \to \infty$. Since s is now a variable, we denote the representation space by $\mathcal{H}_s$ and the holomorphic coherent states by h^s_ζ. The matrices representing the generators J_k will be denoted by S^s_k. By way of motivation, compare the resolution of unity on $\mathcal{H}_s$,

$$\frac{2s+1}{\pi} \int_{\mathbb{C}} d^2\zeta \, (1 + \bar{\zeta}\zeta)^{-2s-2} \, | \, h^s_\zeta \, \rangle \langle \, h^s_\zeta \, | \, = I_s, \tag{1}$$

with that on the representation space $\mathcal{H}$ of $\mathcal{W}$ in terms of the canonical coherent states,

$$\frac{1}{2\pi} \int_{\mathbb{C}} d^2 z \, e^{-\bar{z}z/2} \, | \, \chi_z \, \rangle \langle \, \chi_z \, | \, = I. \tag{2}$$

Note that if we set

$$\zeta = -\frac{z}{2\sqrt{s+1}} \tag{3}$$

and define $\chi_z^s \equiv h_\zeta^s$, then the resolution of unity on $\mathcal{H}_s$ becomes

$$\frac{2s+1}{4\pi(s+1)} \int_{\mathbb{C}} d^2z \, \left(1 + \frac{\bar{z}z}{2(2s+2)}\right)^{-(2s+2)} |\, \chi_z^s \,\rangle\langle\, \chi_z^s \,| = I. \tag{4}$$

If we now take the *formal* limit $s \to \infty$, this coincides with the resolution of unity on $\mathcal{H}$, provided we can show that $\chi_z^s \to \chi_z$. Our task is now (a) to find the *sense* in which this limit is to be taken, (b) to show that the generators S_k^s of $\mathbf{g}$ go over to the generators A, A^* and I of $\mathbf{w}$ and (c) to show that the coherent states χ_z^s go over to the canonical coherent states χ_z.

To properly study the limit $s \to \infty$, we will first of all imbed all the spaces $\mathcal{H}_s$ into $\mathcal{H}$, so that the limit may be considered within $\mathcal{H}$. This is done most easily by using the orthonormal bases obtained by applying S_+^s and A^* to the respective "ground states". An orthonormal basis for $\mathcal{H}$ is given by

$$w_n = (2^n n!)^{-1/2} \, (A^*)^n \chi_0, \qquad n = 0, 1, 2, \ldots, \tag{5}$$

where $A\chi_0 = 0$ and the normalization is determined by the commutation relation $[A, A^*] = 2I$. The generators A and A^* act by

$$\begin{aligned} A w_n &= \sqrt{2n}\, w_{n-1} \\ A^* w_n &= \sqrt{2n+2}\, w_{n+1}. \end{aligned} \tag{6}$$

An orthonormal basis for $\mathcal{H}_s$ is given by

$$w_n^s = \sqrt{\frac{(2s-n)!}{n!(2s)!}} \, \left(S_+^s\right)^n h^s, \qquad n = 0, 1, 2, \ldots, 2s, \tag{7}$$

where $S_-^s h^s = 0$. We imbed $\mathcal{H}_s$ into $\mathcal{H}$ by identifying w_n^s with w_n and defining $S_k^s w_n = 0$ for $n > 2s$. Then for $0 \le n \le 2s$,

$$\begin{aligned} S_-^s w_n &= \sqrt{n(2s-n+1)}\, w_{n-1} \\ S_+^s w_n &= \sqrt{(n+1)(2s-n)}\, w_{n+1} \\ S_3^s w_n &= (n-s) w_n. \end{aligned} \tag{8}$$

To see how the generators S_k^s must be scaled, note that the relation between ζ and z implies that

$$h^s_\zeta \equiv \exp(-\bar\zeta S^s_+)h^s = \exp(\bar z K_+/2)w_0 \equiv \chi^s_z, \tag{9}$$

where

$$K^s_+ = \frac{S^s_+}{\sqrt{s+1}}. \tag{10}$$

Define $K^s_- = S^s_-/\sqrt{s+1}$, so that $K^s_- = \left(K^s_+\right)^*$, and

$$K^s_3 \equiv \frac{1}{2}[K^s_+, K^s_-] = \frac{S^s_3}{s+1}. \tag{11}$$

Then for $0 \le n \le 2s$,

$$\begin{aligned} K^s_- w_n &= \sqrt{\frac{n(2s-n+1)}{s+1}}\, w_{n-1} \\ K^s_+ w_n &= \sqrt{\frac{(n+1)(2s-n)}{s+1}}\, w_{n+1} \\ K^s_3 w_n &= \frac{n-s}{s+1}\, w_n. \end{aligned} \tag{12}$$

If we now take the limit $s \to \infty$ while *keeping n fixed,* we obtain

$$\begin{aligned} K^s_- w_n &\to \sqrt{2n}\, w_{n-1} \\ K^s_+ w_n &\to \sqrt{2n+2}\, w_{n+1} \\ K^s_3 w_n &\to -w_n. \end{aligned} \tag{13}$$

Comparing this with the action of A and A^*, we see that

$$\begin{aligned} K^s_- &\to A \\ K^s_+ &\to A^* \\ K^s_3 &\to -I \end{aligned} \tag{14}$$

as $s \to \infty$ in the *weak operator topology* of $\mathcal{H}$. (These limits are not valid in the strong operator topology since it is necessary to keep n fixed.) Thus we have shown that in the weak sense, the representation of $SU(2)$ goes over to a representation of $\mathcal{W}$ as $s \to \infty$. Note that the operator

$$N^s_- = S^s_3 + s, \tag{15}$$

which satisfies

$$\begin{aligned} N_-^s w_n &= n w_n, \qquad 0 \le n \le 2s, \\ N_-^s w_n &= s w_n, \qquad n > 2s, \end{aligned} \tag{16}$$

has the weak limit

$$\text{w–lim}\, N_-^s \equiv N = \frac{1}{2} A^* A, \tag{17}$$

which is the Hamiltonian for the harmonic oscillator (minus the ground–state energy). If we write $A = X + iP$ with X and P self–adjoint, then

$$N = \frac{1}{2}\left(X^2 + P^2 - I\right). \tag{18}$$

The fact that $K_+^s \to A^*$ means that

$$\chi_z^s \equiv \exp\left(\bar{z} K_+^s/2\right) w_0 \longrightarrow \exp(\bar{z} A^*/2) w_0 \equiv \chi_z, \tag{19}$$

where χ_z are the canonical coherent states and the convergence is in the weak topology of $\mathcal{H}$.

We now examine the limit $s \to \infty$ from a global geometric point of view, using the coherent states. Recall that the expectation vector

$$\tilde{\mathbf{S}}^s(\zeta) \equiv \frac{\langle\, h_\zeta^s \,|\, \mathbf{S}^s\, h_\zeta^s \,\rangle}{\langle\, h_\zeta^s \,|\, h_\zeta^s \,\rangle} \tag{20}$$

ranges over the sphere of radius s centered at the origin, with the north pole corresponding to $\zeta = \infty$. The transformation from S_k^s to K_k^s *deforms* this sphere to an ellipsoid,

$$\frac{|\tilde{K}_+^s(\zeta)|^2}{s+1} + \tilde{K}_3^s(\zeta)^2 = \left(\frac{s}{s+1}\right)^2. \tag{21}$$

When $s \to \infty$, this ellipsoid splits into the two planes $\tilde{K}_3^s \to \pm 1$. Our weak limit $K_3^s \to -I$ only picked out the lower plane. We could have picked out the upper plane by imbedding $\mathcal{H}_s$ into $\mathcal{H}$ differently, namely by identifying w_{2s-n} with w_n for $n = 0, 1, 2, \ldots, 2s$. In that case we would have obtained the weak limits

$$\begin{aligned} K^s_- &\to A^* \\ K^s_+ &\to A \\ K^s_3 &\to I. \end{aligned} \tag{22}$$

In terms of the coherent states, this corresponds to using the highest–weight vector instead of the lowest–weight vector or, equivalently, using a chart centered about the north pole rather than the south pole, for example, by using as reference point the element $g_1 = e^{-i\pi J_2}$. The corresponding harmonic–oscillator Hamiltonian is the weak limit of the operator

$$N^s_+ \equiv s - S^s_3 = 2s - N^s_-. \tag{23}$$

The expectation values of N^s_- and N^s_+,

$$\begin{aligned} \tilde{N}^s_-(\zeta) &= \frac{2s\bar{\zeta}\zeta}{1+\bar{\zeta}\zeta} \\ \tilde{N}^s_+(\zeta) &= \frac{2s}{1+\bar{\zeta}\zeta}, \end{aligned} \tag{24}$$

are related by *inversion:*

$$\tilde{N}^s_+(\zeta) = \tilde{N}^s_-(1/\zeta). \tag{25}$$

The splitting of the ellipsoid into the two planes can be understood from the point of view of representation theory by writing the irreducibility condition in terms of the K's:

$$\frac{1}{2s+2}\left(K^s_+K^s_- + K^s_-K^s_+\right) + \left(K^s_3\right)^2 = \frac{s}{s+1}I_s. \tag{26}$$

When $s \to \infty$, this implies formally that $\left(K^s_3\right)^2 \to I$. The subspaces $\mathcal{H}_\pm$ on which $K^s_3 \to \pm I$ are invariant in the limit, hence the limiting representation of $\mathcal{W}$ is *reducible*. Evidently, by taking the limits in the weak topology we are able to pick out just one irreducible component at a time.

There is an interesting analogy between the above analysis and the non–relativistic limit of Klein–Gordon theory, which we now point out for those readers already familiar with the latter. (The non–relativistic limit will be discussed in the next chapter.) The relation $S^s_3/(s+1) \to \pm I$ corresponds to $P_0/mc^2 \to \pm I$ in the limit $mc^2 \to \infty$,

where P_0 is the relativistic energy operator. As will be seen in the next chapter, the Poincaré group contracts, in this approximation, to a semidirect product of the 7–dimensional Weyl–Heisenberg group and the rotation group. A first–order correction is obtained by expanding

$$P_0 = \pm\sqrt{m^2c^4 + \mathbf{P}^2c^2} \sim \pm\left(mc^2 + \mathbf{P}^2/2m\right) \equiv \pm\left(mc^2 + H\right), \quad (27)$$

where H is the non–relativistic free Schrödinger Hamiltonian. Similarly, it follows from

$$\begin{aligned} &\frac{1}{2}\left(S^s_+S^s_- + S^s_-S^s_+\right) + \left(S^s_3\right)^2 = s(s+1) \\ &\frac{1}{2}\left(S^s_+S^s_- - S^s_-S^s_+\right) - S^s_3 = 0 \end{aligned} \quad (28)$$

that

$$\left(S^s_3 - \tfrac{1}{2}\right)^2 = \left(s + \tfrac{1}{2}\right)^2 - S^s_+S^s_-, \quad (29)$$

hence formally

$$\begin{aligned} S^s_3 - \tfrac{1}{2} &= \pm\sqrt{\left(s+\tfrac{1}{2}\right)^2 - S^s_+S^s_-} \\ &\sim \pm\left(s + \tfrac{1}{2}\right) \mp \frac{S^s_+S^s_-}{2s+1}, \end{aligned} \quad (30)$$

from which

$$\begin{aligned} S^{s(-)}_3 &\sim -s + \frac{1}{2}K^s_+K^s_-, \\ S^{s(+)}_3 &\sim s + 1 - \frac{1}{2}K^s_+K^s_-, \end{aligned} \quad (31)$$

which corresponds to $P_0^{(\pm)} \sim \pm(mc^2 + H)$. The analogy between the large–spin and the non–relativistic limits can be summarized as follows: The Poincaré group corresponds to $SU(2)$, the energy P_0 to S^s_3, the rest energy mc^2 to $-s$, and the non–relativistic Hamiltonian H to the harmonic oscillator hamiltonian N. The analog of the central extension of the Galilean group (which is the contraction limit of the Poincaré group, as discussed in the next chapter) is the *Oscillator group* (Streater [1967]), whose generators are A, A^*, I and N, with the commutation relations

$$[N, A] = -A, \qquad [N, A^*] = A^*. \tag{32}$$

Finally, we note that the above analysis explains a well–known relationship between the harmonic oscillator and the canonical coherent states, namely that the latter evolve naturally under the harmonic oscillator dynamics. We first derive the corresponding relation within $SU(2)$, i.e. for finite s. Consider the behavior of h^s_ζ under the "evolution operator" $\exp(-itN^s_-)$. We wish to express

$$e^{-itN^s_-}\chi^s_z = e^{-its}e^{-itS_3}e^{-\bar\zeta S_+}h^s \tag{33}$$

in terms of the coherent states χ^s_z, hence we need to display the operator on the right in the reverse Gaussian form

$$\exp(-\bar\zeta' S^s_+)\exp(2d'S_3)\exp(\xi' S_-). \tag{34}$$

As explained earlier, it suffices to do the computation on 2×2 matrices:

$$\begin{aligned} e^{-itJ_3}e^{-\bar\zeta J_+} &= \begin{pmatrix} e^{-it/2} & 0 \\ 0 & e^{it/2}\end{pmatrix}\begin{pmatrix} 1 & -\bar\zeta \\ 0 & 1\end{pmatrix} \\ &\equiv \begin{pmatrix} 1 & -\bar\zeta' \\ 0 & 1\end{pmatrix}\begin{pmatrix} e^{d'} & 0 \\ 0 & e^{-d'}\end{pmatrix}\begin{pmatrix} 1 & 0 \\ \xi' & 1\end{pmatrix}. \end{aligned} \tag{35}$$

This gives $d' = -it/2$, $\zeta' = e^{it}\zeta$ and $\xi' = 0$. Hence

$$\begin{aligned} e^{-itN^s}\chi^s_z &= e^{-its}e^{-\bar\zeta' S^s_+}e^{-itS^s_3}h \\ &= e^{-\bar\zeta' S^s_+}h = h_{\zeta'} \\ &= \chi^s_{z(t)}, \end{aligned} \tag{36}$$

where $z(t) = e^{it}z$. (This is intuitively obvious, since $\exp(-itS^s_3)$ rotates the 1–2 plane clockwise by an angle t, hence it rotates the coordinate z counterclockwise by an angle t.) In the limit $s \to \infty$, this gives the well–known result (Henley and Thirring [1962]) that the *set* of canonical coherent states is invariant under the harmonic oscillator time evolution, with individual coherent states moving along the classical trajectories $z(t)$ determined by the initial conditions $z = x - ip$ in phase space. The above shows that the same is true within $SU(2)$, i.e. for finite s, where it is a consequence of the fact that N^s_- is essentially the generator of rotations about the 3–axis.

Note: After this section was written, I learned from R. F. Streater that a related construction was made by Dyson [1956].

Chapter 4
COMPLEX SPACETIME

4.1. Introduction

Relativistic quantum mechanics is a synthesis of special relativity and ordinary (i.e., non-relativistic) quantum mechanics. The former is based on the Lorentzian geometry of spacetime, while the latter is usually obtained from classical mechanics by a somewhat mysterious set of rules known as "quantization" in which classical observables, which are functions on phase space, suddenly become operators on a Hilbert space. Classical mechanics, in turn, can be formulated in terms of Newtonian space-time (the Lagrangian approach) or it can be based on the symplectic geometry of phase space (see Abraham and Marsden [1978]). The latter, called the *Hamiltonian approach*, is usually considered to be deeper and more powerful, and its study has virtually turned modern classical mechanics into a branch of differential geometry. Yet, the standard formalism of relativistic quantum mechanics rests solely on the geometry of spacetime. Symplectic geometry, so prominent in classical mechanics, seems to have disappeared without a trace.

In this chapter we develop a formulation of relativistic quantum mechanics in which symplectic geometry plays an important role. This will be done by studying the role of phase space in relativity and discovering its counterpart in relation to the Poincaré group, which is the invariance group of Minkowskian spacetime. It turns out that the Perelomov construction fails for relativistic particles (the physically relevant representations are not square-integrable), and an alternative route must be taken. The result is a formalism based on *complex spacetime* which, we show, may be regarded as a relativistic extension of classical phase space. As a by-product, two long-standing inconsistencies of relativistic quantum mechanics (in its standard spacetime formulation) are resolved, namely the problems of *localization* and *covariant probabilistic interpretation*. Rather than being *sharply* localizable in *space* (which leads to conflicts with causality; see Newton

and Wigner [1949] and Hegerfeldt [1985]), particles in the new formulation are at best *softly* localizable in *phase space*. This is just a covariant version of the situation in the coherent–state representation. But whereas for non-relativistic particles both the Schrödinger representation and the coherent–state representation give equally consistent theories, the spacetime formulation of relativistic quantum mechanics is inconsistent because it lacks a genuine probabilistic interpretation, a situation remedied by the phase-space formulation (section 4.5).

4.2. Relativity, Phase Space and Quantization

At first it appears that phase space and spacetime are mutually exclusive: The phase-space coordinates of a particle are in one-to-one correspondence with the initial conditions which determine its classical motion, i.e. its worldline. Hence the phase space can be identified with either the set of all initial conditions or the set of worldlines of the classical particle. In either case, time is treated differently from space: For initial conditions, an arbitrary "initial" time is chosen; for world-lines, the dynamical "flow" is factored out. Furthermore, *locality* in spacetime is lost in either case.

In this section we confine ourselves to the physical case $s = 3$, i.e., three–dimensional space. Our approach will be to leave spacetime intact and, instead, consider its group of symmetries, the *Poincaré group*, which is defined as follows: Let u be a four-vector. We denote its time component (with respect to an arbitrary reference frame) by u^0 and its space components by $\mathbf{u} = (u^1, u^2, u^3)$. The invariant Lorentzian inner product of two such vectors is defined as

$$(u, v) \equiv uv \equiv c^2 u^0 v^0 - \mathbf{u} \cdot \mathbf{v} \equiv g_{\mu\nu} u^\mu v^\nu. \tag{1}$$

Later we will set $c = 1$ (which amounts to measuring time as the distance traveled by light), but for the present it is important to include c, since the non-relativistic limit $c \to \infty$ will be considered.

The *Lorentz group* $\mathcal{L}$ is the set of all linear transformations $\Lambda : \mathbb{R}^4 \to \mathbb{R}^4$ which leave the inner product invariant:

$$(\Lambda u, \Lambda v) = (u, v) \quad \forall u, v \in \mathbb{R}^4. \tag{2}$$

$\mathcal{L}$ includes transformations which reverse the orientation of time and ones which reverse the orientation of space. Such space– and time reflections split $\mathcal{L}$ into four connected components. The component

connected to the identity (whose elements reverse neither the orientations of time nor of space) is called the *restricted Lorentz group* and denoted by $\mathcal{L}_0$. If we let $u^4 \equiv cu^0$, so that $uv = -\mathbf{u}\cdot\mathbf{v} + u^4v^4$, we can identify $\mathcal{L}_0$ with $SO(3,1)_+$ (the plus sign indicates that the orientation of time, hence also of space, is preserved separately.) The *Poincaré group* $\mathcal{P}$ is defined as the set of all Lorentz transformations combined with spacetime translations:

$$\mathcal{P} = \{(a,\Lambda)|\Lambda \in \mathcal{L}, a \in \mathbb{R}^4\} \tag{3}$$

where (a,Λ) acts on $\mathbb{R}^4$ by

$$(a,\Lambda)u = \Lambda u + a. \tag{4}$$

We will be dealing with the *restricted Poincaré group* $\mathcal{P}_0$ where $\Lambda \in \mathcal{L}_0$. The reason for our interest in $\mathcal{P}_0$ is that it parametrizes all reference frames which can be obtained from a given reference frame by a continuous motion. (By "motion" we mean any spacetime translation, rotation or boost, including physically impossible "motions" such as space–like translations and translations backwards in time.) Now to specify a reference frame (a,Λ) relative to some fixed reference frame (located, say, at the origin in spacetime), we must give its *origin* (namely, a), its *velocity* and its *spatial orientation* (all relative to the fixed frame). Thus if we *ignore* the spatial orientation by factoring out the rotation subgroup $SO(3)$ of $\mathcal{L}_+^\uparrow$, the resulting seven-dimensional homogeneous space

$$\mathcal{C} \equiv \mathcal{P}_0/SO(3) \tag{5}$$

can be identified with the set of positions in spacetime (*events*) together with all possible velocities at these events. The set of all (future-pointing) four-velocities is a hyperboloid

$$\Omega_c^+ \equiv \{u \in \mathbb{R}^4|\ u^2 \equiv (u,u) = c^2,\ u^0 > 0\}. \tag{6}$$

Thus

$$\mathcal{C} \approx \mathbb{R}^4 \times \Omega_c^+, \tag{7}$$

which is an extension of classical phase space obtained by including time along with the space coordinates. Such an object is usually called a *state space*. Strictly speaking, $\mathcal{C}$ is an extended *velocity* phase space rather than an extended momentum phase space; this appears to be

the price for retaining locality in spacetime. What matters for us is that it will have the required symplectic (more precisely, *contact*) structure. From its construction, it is clear that $\mathcal{C}$ is a relativistically covariant object; it is not *in*variant since the choice of $SO(3)$ in $\mathcal{P}_0$ is frame-dependent. We will see that $\mathcal{C}$ combines the geometries of spacetime and phase space in a natural way.

Thus, rather than conflicting with relativity, the concept of phase space actually *follows* from it: Appending time to the geometry of space means appending velocity-changing transformations (which are just rotations in a space-time plane) to the group of rigid motions, hence the enlarged group contains velocity coordinates in addition to space coordinates. In a sense, $\mathcal{P}_0$ itself is actually a "super phase space" since it furthermore contains information on the spatial orientation, which is needed to include *spin* degrees of freedom along with the translational degrees of freedom. $\mathcal{P}_0$ even has a natural generalization to the case of *curved* spacetime, namely the *frame bundle* of all "orthonormal" frames (with respect to the given curved metric) at all possible events. (For the definition and study of frame bundles, see Kobayashi and Nomizu [1963, 1969].)

In our review of canonical coherent states (chapter 1), we saw that the classical phase space resulted from the Weyl-Heisenberg group $\mathcal{W}$, which, unlike $\mathcal{P}_0$ was not a symmetry group of the theory but merely a Lie group generated by the fundamental dynamical observables of position and momentum at a fixed time. We will now see that $\mathcal{W}$ is related to the non-relativistic limit of $\mathcal{P}_0$ in two distinct ways: as a normal subgroup, and as a homogeneous space. This insight will play a key role in generalizing the canonical coherent states to the relativistic case. It turns out that the role of $\mathcal{W}$ as a *group* has no relativistic counterpart, whereas its role as a homogeneous space does: its relativistic generalization is $\mathcal{C}$.

Consider the Lie algebra $\wp$ of $\mathcal{P}_0$, which is spanned by the generators P_k of spatial translations, P_0 of time translations, J_k of spatial rotations and K_k of pure Lorentz transformations (the "boosts"). The Lie brackets of $\wp$ are given by

$$\begin{array}{ll} [J_j, J_k] = iJ_l & [J_j, K_k] = iK_l \\ [P_0, K_r] = iP_r & [J_j, P_k] = iP_l \\ [K_j, K_k] = -ic^{-2}J_l & [P_r, K_s] = ic^{-2}\delta_{rs}P_0 \end{array} \tag{8}$$

where (j, k, l) is a cyclic permutation of (1,2,3), $r, s = 1, 2, 3$ and all unspecified brackets vanish. The physical dimensions of these generators are as follows: P_0 is a reciprocal time, P_k is a reciprocal lenght,

J_k is dimensionless (reciprocal angle) and K_k is a reciprocal velocity.

Notice that so far nothing has been said about quantum mechanics. $\wp$ is simply the Lie algebra of the infinitesimal motions of Minkowskian spacetime, a classical concept. (The unit imaginary i in eq (9) can be removed by replacing P_μ with $-iP_\mu$, J_k with $-iJ_k$ and K_k with $-iK_k$; i is included because we anticipate that in quantum mechanics these generators become self-adjoint operators.) Quantum mechanics is now introduced through the following postulate:

(Q). The formalism of relativistic quantum theory is based on a unitary (though possibly reducible) representation of $\mathcal{P}_0$.

That is, the representation provides the quantum–mechanical Hilbert space, and the generators of $\wp$, which by unitarity are represented by self-adjoint operators, are interpreted as the basic physical observables: P_0 as the energy, P_k as the momentum and J_k as the angular momentum. One may then consider perturbations by introducing interactions or gauge fields. In fact, the assumption *(Q)* is very general in scope; it serves as one of the axioms in axiomatic quantum field theory (Streater and Wightman [1964]). Unlike the usual prescription of "quantization" in non-relativistic quantum mechanics, *(Q)* is both mathematically and physically unambiguous. Yet, *(Q) implies and, at the same time, supercedes the canonical commutation relations!* To see this, consider the non-relativistic limit $c \to \infty$ of $\wp$. Letting

$$c^{-2}P_0 \equiv M, \tag{9}$$

we see that in the limit $c \to \infty$, $\wp$ "contracts" to a Lie algebra $\mathbf{g_1}$ with generators M, P_k, J_k, K_k satisfying

$$\begin{array}{ll} [J_j, J_k] = iJ_l & [J_j, K_k] = iK_l \\ [M, K_r] = 0 & [J_j, P_k] = iP_l \\ [K_j, K_k] = 0 & [P_r, K_s] = i\delta_{rs}M \end{array} \tag{10}$$

and all other brackets vanishing. Note that (a) M is a central element of $\mathbf{g_1}$ and (b) M, P_k and K_k span an invariant subalgebra $\mathbf{w}$ of $\mathbf{g_1}$ which is isomorphic to the Weyl–Heisenberg algebra, with M playing the role of the central element E. Hence if $\mathcal{G}_1$ denotes the connected, simply connected Lie group generated by $\mathbf{g_1}$, then the invariant subgroup of $\mathcal{G}_1$ generated by $\mathbf{w}$ is isomorphic to the Weyl–Heisenberg group $\mathcal{W}$. The remaining generators J_k of $\mathbf{g_1}$ span the Lie algebra $so(3)$ of the spatial rotation group $SO(3)$, so $\mathcal{G}_1$ is the semi-direct product of $\mathcal{W}$ with $SO(3)$:

$$\mathcal{G}_1 = \mathcal{W} \circledS SO(3). \tag{11}$$

Now suppose that the unitary representation of $\mathcal{P}_+^\uparrow$ in assumption *(Q) is irreducible.* [Assumption *(Q)* means that we are dealing with a general quantum *system*, possibly a system of interacting particles or even quantum fields; it is the additional assumption of irreducibility which makes this system *elementary*, roughly a single *particle*. Hence the concept of *position*, discussed below, is only now admissible.] Assuming that the formal limit $c \to \infty$ of Lie algebras rigorously induces a corresponding limit at the representation level (and this is indeed the case, as will be shown later), assumption *(Q)* implies that we have a unitary irreducible representation of $\mathcal{G}_1$ in that limit. Then M, P_k and K_k are represented by self-adjoint operators on some Hilbert space, which we will denote by the same symbols. Irreducibility implies that the central element M has the form mI, where m is a real number and I denotes the identity operator. Assume $m > 0$ (this is physically necessary since m will be interpreted as a *mass*) and let

$$X_k = -(1/m)K_k. \tag{12}$$

Then eq. (10) shows that X_k and P_k satisfy the canonical commutation relations:

$$[X_r, P_s] = i\delta_{rs}I, \tag{13}$$

thus X_k behaves like a position operator. This shows that the assumption *(Q)*, which is conceptually simple, mathematically precise, relativistically invariant and very general, actually *implies* the much less satisfactory "quantization" prescription in the non-relativistic limit, under the additional assumption of irreducibility.

How does it happen that classical relativistic geometry, as represented by $\mathcal{P}_0$, when combined with assumption *(Q)*, yields the mysterious canonical commutation relations? To understand this, note that eq. (10) came from the relativistic Lie bracket

$$[P_r, K_s] = ic^{-2}\delta_{rs}P_0 \tag{14}$$

which states that boosting (accelerating) in any given spatial direction does not commute with translating in the same direction. This, in turn, is a consequence of the fact that Einsteinian space is not absolute since in the accelerated frame there is a (Lorentz) contraction in the direction of motion, so first translating and then boosting is not the same as first boosting and then translating. In the non-relativistic limit this gives the canonical commutation relations. No such easy

derivation of these relations would have been possible without invoking Relativity. For had we begun with Newtonian space-time, the appropriate invariance group would have been not $\mathcal{P}_0$ but the Galilean group $\mathcal{G}$. Since Newtonian space is absolute and hence unaffected by boosting to a moving frame, the Galilean boosts ${K_k}'$ commute with with the Galilean generators of translations ${P_k}'$, hence yield no canonical commutation relations and no associated uncertainty principle. In the case of $\mathcal{G}_1$, the canonical commutation relations are a remnant of relativistic invariance. Thus *the uncertainty principle originates, in some sense, in "classical" Relativity theory!*

Eq. (12) states that an acceptable set of position operators for a non-relativistic particle is given in terms of the generators of Galilean boosts (more precisely, the boosts of a central extension of the Galilean group, as explained below). It is interesting to see how this comes about from a more intuitive, physical point of view, since position operators are problematic in relativistic quantum mechanics (as will be discussed later) but the boosts have natural relativistic counterparts. To gain insight, we will now give two additional rough but intuitive arguments for the validity of eq. (12).

1. For a spinless particle, the generators of the Poincaré group can be realized as operators on a space of functions over spacetime (namely, the space of solutions of the Klein–Gordon equation) by

$$\begin{aligned} P_0 &= i\frac{\partial}{\partial x_0} \\ P_k &= -i\frac{\partial}{\partial x_k} \\ J_k &= x_l P_m - x_m P_l \\ K_k &= x_0 P_k - c^{-2} x_k P_0 \end{aligned} \tag{15}$$

where (k, l, m) is a cyclic permutation of $(1, 2, 3)$. In the non-relativistic limit, $P_0 \to mc^2$, so

$$-(1/m)K_k \to x_k - x_0 P_k/m, \tag{16}$$

which displays $-(1/m)K_k$ as the operator of multiplication by x_k (the usual non-relativistic position operator) minus the distance the particle has traveled in time $t = x_0$ while going at a velocity $\mathbf{v} = \mathbf{P}/m$. This is just the *initial* position of the particle at time $t = 0$; that is, the non-relativistic limit of $-(1/m)K_k$ is just the (non-relativistic) position operator in the *Schrödinger picture*, where operators are constant

while states vary with time.

2. The second explanation of eq. (12) begins with the relativistic state space $\mathcal{C} \approx \mathbb{R}^4 \times \Omega_c^+$ and considers the non-relativistic limit. As $c \to \infty$, the hyperboloid Ω_c^+ flattens out to a (three-dimensional) hyperplane at infinity, the non-relativistic *velocity space* $\mathcal{V} \approx \mathbb{R}^3$. In that limit, the boosts K_k commute and become the generators of translations in $\mathcal{V}$. If the cartesian coordinates of $\mathcal{V}$ are v_k, then

$$K_k \to -i\frac{\partial}{\partial v_k}. \tag{17}$$

Now the velocities v_k are related to the momenta p_k by $v_k = p_k/m$, where m is the mass. Thus in the limit

$$K_k = -mi\frac{\partial}{\partial p_k}. \tag{18}$$

But in the *momentum representation* of non-relativistic quantum mechanics , the position operators are represented by

$$X_k = i\frac{\partial}{\partial p_k}, \tag{19}$$

which again gives agreement with eq. (12).

Now that we have an acceptable relativistic generalization of the classical phase space, let us return to our goal of extending the canonical coherent–state representation to relativistic particles. We have seen that the Weyl–Heisenberg group, on which this representation is based, is isomorphic to an invariant subgroup of the non-relativistic limit of $\mathcal{P}_+^\uparrow$. However, this subgroup is *not* the non-relativistic limit of any subgroup of $\mathcal{P}_0$, since the Lie brackets of K_k, P_k and P_0 do not close due to

$$[K_j, K_k] = -ic^{-2}J_l, \ (j,k,l \text{ cyclic }). \tag{20}$$

That is, we cannot simply generalize the canonical coherent states by choosing the right subgroup of $\mathcal{P}_0$. However, eq. (11) shows that $\mathcal{W}$ also plays another role in the group $\mathcal{G}_1$, namely as a *homogeneous space*:

$$\mathcal{W} = \mathcal{G}_1/SO(3). \tag{21}$$

In this form, it does have an obvious relativistic counterpart, namely $\mathcal{C}$. In retrospect, the role of $\mathcal{W}$ as a group is purely incidental to quantum mechanics , since there is no fundamental reason why the set of all dynamical states should form a group. On the other hand, this set should certainly be a homogeneous space under the group of motions, since this group must transform dynamical states into one another and this action can be assumed to be transitive (otherwise we may as well restrict ourselves to an orbit). In either of its roles relative to $\mathcal{G}_1$, $\mathcal{W}$ acquires a slightly different physical interpretation from the one it had in relation to the canonical commutation relations: Since the group–manifold of $\mathcal{W}$ is generated by the vector fields P_k, K_k and M, the coordinates on $\mathcal{W}$ are the positions x_k (generated by P_k), the velocities v_k (generated by K_k) and the variable generated by M, which is a degenerate form of "time" inherited form Relativity through the limit $c^{-2}P_0 \to M$. By comparison, the coordinates on the original Weyl–Heisenberg group were x_k (generated by P_k), the momentum p_k (generated by X_k) and a dimensionless "phase angle" ϕ generated by the identity operator. With this new interpretation, $\mathcal{W}$ is the product of velocity phase space with "time" and truly represents the non-relativistic limit of $\mathcal{C}$.

We will construct a coherent-state representation of $\mathcal{P}_0$ by first discovering its non-relativistic limit. This limit will be a representation of a *quantum mechanical* version $\mathcal{G}_2$ of the Galilean group $\mathcal{G}$ and will be seen to be a close relative of the canonical coherent–state representation of $\mathcal{W}$. It is therefore necessary first to understand just how the group $\mathcal{G}$ is related to the Poincaré group $\mathcal{P}_0$ and its non–relativistic limit $\mathcal{G}_1$. Again, we will do everything at the level of Lie algebras. There is no problem with globalization.

The universal enveloping algebra of $\wp$ contains the mass-squared operator

$$M^2 = c^{-4}{P_0}^2 - c^{-2}({P_1}^2 + {P_2}^2 + {P_3}^2) = c^{-4}{P_0}^2 - c^{-2}\mathbf{P}^2, \qquad (22)$$

which is a Casimir operator, i.e. commutes with all generators in $\wp$. Assuming that both P_0 and M are represented by positive operators and that M is invertible (and this will be the case for massive particles), we have formally for large c:

$$c^{-2}P_0 = (M^2 + c^{-2}\mathbf{P}^2)^{1/2} = M + \mathbf{P}^2/2Mc^2 + O(c^{-4}), \qquad (23)$$

where we have used the fact that M commutes with P_k. The operator

$$H = \mathbf{P}^2/2M \tag{24}$$

is just the Hamiltonian for the non-relativistic free particle, hence generates its time translations and must be included in the Lie algebra of the Galilean group. Let us therefore try to append it to the Lie algebra $\mathbf{g}_1$ of $\mathcal{G}_1$. Indeed, by eq. (10),

$$\begin{array}{ll} [H, P_k] = 0 & [H, M] = 0 \\ [H, J_k] = 0 & [H, K_k] = iP_k, \end{array} \tag{25}$$

showing that

$$\mathbf{g}_2 \equiv \text{ Span } \{K_k, P_k, J_k, M, H\} \tag{26}$$

actually forms a Lie algebra with Lie brackets given by eqs. (10) and (25). Clearly $\mathbf{g}_2$ contains $\mathbf{g}_1$ as a subalgebra. We will refer to the corresponding Lie group $\mathcal{G}_2$ as the *quantum mechanical Galilean group*. It is the group of translations, rotations, boosts, dynamics (i.e., time translations) and multiplications by constant phase factors (generated by M) acting on the wave functions of a non-relativistic quantum mechanical particle. The relation of $\mathcal{G}_2$ to the *classical* Galilean group $\mathcal{G}$ is as follows: The subgroup generated by M, which can be identified with the group of real numbers $\mathbb{R}$, is central; then $\mathcal{G}$ is obtained from $\mathcal{G}_2$ by factoring out $\mathbb{R}$:

$$\mathcal{G} = \mathcal{G}_2/\mathbb{R}. \tag{27}$$

The action of $\mathbb{R}$ on quantum mechanical wave functions, which amounts to a multiplication by a constant phase factor, is a necessary part of $\mathcal{G}_2$ because of $[P_r, K_s] = i\delta_{rs}M$ which, as we have seen, is related to the uncertainty principle. Factoring out this action means ignoring that phase factor, so it is reasonable that it should give the classical Galilean group. On the Lie algebra level, it amounts to setting $M = 0$. Had we included Planck's constant $\hbar$ in eq. (10), this would have amounted to taking the classical limit $\hbar \to 0$. The above relation between $\mathbb{R}$, $\mathcal{G}_2$ and $\mathcal{G}$ in an example of a *central extension* (Varadarajan [1969]). One says that $\mathcal{G}_2$ is a central extension of $\mathbb{R}$ by $\mathcal{G}$.

The fact that $\mathcal{W}$ is a subgroup of $\mathcal{G}_2$ was noted by Bargmann [1954]; that representations of $\mathcal{G}_2$ are contractions of representations of the Poincaré group was shown by Mackey [1955].*

* I thank R. F. Streater for these remarks.

4.3. Galilean Frames

Our object in this section is to construct coherent states which are naturally associated with free non–relativistic particles, just as the canonical coherent states are associated with the Weyl–Heisenberg group or the Harmonic oscillator and the spin coherent states are associated with with $SU(2)$. An obvious starting point would be to apply the Klauder–Perelomov method (chapter 1) to the quantum–mechanical Galilean group $\mathcal{G}_2$, since it is this group which describes such particles. However, this method fails, due to the fact that all the representations of physical interest are not square–integrable. Therefore we will follow a more pedestrian route. Our main guides will be analyticity (which, it turns out, follows from an important physical condition) and "physical intuition."

We return to the general case where the configuration space is $\mathbb{R}^s$ instead of $\mathbb{R}^3$. For simplicity, we restrict ourselves to spinless particles. It is not difficult to include spin, as will be shown later. The states of such particles are described by complex–valued wave functions $f(\mathbf{x}, t)$ of position $\mathbf{x}$ and time t which are are square–integrable with respect to $\mathbf{x}$ at any time t. Their evolution in time is given by the Schrödinger equation

$$i\frac{\partial f}{\partial t} = Hf, \tag{1}$$

where

$$H = -\frac{1}{2m}\Delta \tag{2}$$

is the Hamiltonian operator, and Δ is the Laplacian acting on $L^2(\mathbb{R}^s)$. Since H is self–adjoint, though unbounded, the solutions are given through the unitary one–parameter group $U(t) = \exp(-itH)$:

$$\begin{aligned} f(\mathbf{x}, t) &= (U(t)f)(\mathbf{x}) \\ &= (2\pi)^{-s} \int_{\mathbb{R}^s} d^s\mathbf{p}\, \exp[-it\mathbf{p}^2/2m + i\mathbf{p}\cdot\mathbf{x}]\hat{f}(\mathbf{p}), \end{aligned} \tag{3}$$

where it is assumed that $f(\mathbf{x}, 0)$, hence also its Fourier transform $\hat{f}(\mathbf{p})$, is in $L^2(\mathbb{R}^s)$.

The key to our approach will be to note that H is a non–negative operator, hence the evolution group $U(t)$ can be analytically continued to the lower–half complex time plane $\mathbb{C}^-$ as

$$U(t-iu) = \exp[-i(t-iu)H] = e^{-itH}e^{-uH}, \quad u > 0. \tag{4}$$

(Note also that since H is unbounded, no analytic continuation to the upper–half time plane is possible.) The operator e^{-uH} is familiar from two other contexts: it constitutes the evolution semigroup for the heat equation (where u is time), and it is also the unnormalized density matrix for the Gibbs canonical ensemble describing the statistical equilibrium of a quantum system at temperature T (where $u = 1/kT$). To get a feel for our use of this operator, let us be heuristic for a moment and consider what happens when a free *classical* free particle of mass m is evolved in complex time $\tau = t - iu$. If its initial position and momentum are $\mathbf{x}$ and $\mathbf{p}$ respectively, then its new position will be

$$\begin{aligned} \mathbf{z}(\tau) &= \mathbf{x} + (t-iu)\mathbf{p}/m \\ &= (\mathbf{x} + t\mathbf{p}/m) - i(u/m)\mathbf{p} \\ &= \mathbf{x}(t) - i(u/m)\mathbf{p}. \end{aligned} \tag{5}$$

Since $\mathbf{x}(t)$ is just the position evolved in real time t, we see that $\mathbf{z}(t)$ is, in fact, a *complex phase space coordinate* of the same type we encountered in the construction of the canonical coherent states! Armed with this intuition, let us now return to quantum mechanics and see if this idea has a quantum mechanical counterpart. The operator e^{-uH}, when applied to any function in $L^2(\mathbb{R}^s)$, gives

$$\begin{aligned} f_u(\mathbf{x}) &\equiv (e^{-uH}f)(\mathbf{x}) \\ &= (2\pi)^{-s} \int_{\mathbb{R}^s} d^s\mathbf{p}\, \exp[-u\mathbf{p}^2/2m + i\mathbf{p}\cdot\mathbf{x}]\hat{f}(\mathbf{p}). \end{aligned} \tag{6}$$

If we replace $\mathbf{x}$ in the integrand by an arbitrary $\mathbf{z} \in \mathbb{C}^s$, the integral still converges absolutely since the quadratic term in the exponent dominates the linear term for large $|\mathbf{p}|$. Clearly the resulting function is *entire* in $\mathbf{z}$ (differentiating the integrand with respect to z_k still gives an absolutely convergent integral). This shows that the group of Galilean space–time translations,

$$U(\mathbf{x}, t) = \exp\bigl(-itH + i\mathbf{x}\cdot\mathbf{P}\bigr), \tag{7}$$

extends analytically to a *semigroup* of *complex* space–time translations

$$U(\bar{\mathbf{z}}, \bar{\tau}) = \exp\bigl(-i\bar{\tau}H + i\bar{\mathbf{z}} \cdot \mathbf{P}\bigr) \tag{8}$$

defined over the complex space–time domain

$$\mathcal{D} = \{\, (\mathbf{z}, \tau) \,|\, \mathbf{z} \in \mathbb{C}^s, \tau \in \mathbb{C}^- \,\}. \tag{9}$$

This translation semigroup can be combined with the rotations and boosts to give an analytic semigroup $\mathcal{G}_2^c$ extending $\mathcal{G}_2$.

Let $\mathcal{H}_u$ be the vector space of all the entire functions $f_u(\mathbf{z})$ as $\hat{f}(\mathbf{p})$ runs through $L^2(\mathbb{R}^s)$. Then

$$f_u(\mathbf{z}) = \langle\, e^u_{\mathbf{z}} \,|\, \hat{f}\,\rangle, \tag{10}$$

where

$$\begin{aligned} e^u_{\mathbf{z}}(\mathbf{p}) &= (2\pi)^{-s} \exp[-u\mathbf{p}^2/2m - i\mathbf{p} \cdot \bar{\mathbf{z}}] \\ &= (2\pi)^{-s} \exp\left[m\mathbf{y}^2/2u - \frac{u}{2m}(\mathbf{p} - m\mathbf{y}/u)^2 - i\mathbf{p} \cdot \mathbf{x}\right] \end{aligned} \tag{11}$$

are seen to be Gaussian wave packets in momentum space with expected position and momentum given in terms of $\mathbf{z} \equiv \mathbf{x} - i\mathbf{y}$ by

$$\langle\, X_k \,\rangle = x_k \quad \text{and} \quad \langle\, P_k \,\rangle = (m/u)y_k. \tag{12}$$

The $e^u_{\mathbf{z}}$'s are easily shown to have minimal uncertainty products. The momentum uncertainty can be read off directly from the exponent and is

$$\Delta_{P_k} = \sqrt{m/2u}, \tag{13}$$

hence

$$\Delta_{X_k} = \sqrt{u/2m}. \tag{14}$$

We now have our prospective coherent states and their label space $M = \mathbb{C}^s$. To construct a coherent–state representation, we need a measure on M which will make the $e^u_{\mathbf{z}}$'s into a frame. Since the $e^u_{\mathbf{z}}$'s are Gaussian, the measure in not difficult to find:

$$d\mu_u(\mathbf{z}) = (m/\pi u)^{s/2} \exp\left(-m\mathbf{y}^2/u\right) d^s\mathbf{x}\, d^s\mathbf{y}. \tag{15}$$

Defining

$$\langle\, f \,|\, g \,\rangle_{\mathcal{H}_u} \equiv \langle\, f_u \,|\, g_u \,\rangle \equiv \int d\mu_u(\mathbf{z}) \overline{f_u(\mathbf{z})} g_u(\mathbf{z}), \tag{16}$$

we have

Theorem 4.1.

(a) $\langle \cdot \,|\, \cdot \rangle_{\mathcal{H}_u}$ *is an inner product on* $\mathcal{H}_u$ *under which* $\mathcal{H}_u$ *is a Hilbert space.*

(b) The map e^{-uH} *is unitary from* $L^2(\mathbb{R}^s)$ *onto* $\mathcal{H}_u$.

(c) The $e^u_{\mathbf{z}}$*'s define a resolution of unity on* $L^2(\mathbb{R}^s)$ *given by*

$$\int_{\mathbb{C}^s} d\mu_u(\mathbf{z}) \;|\, e^u_{\mathbf{z}} \,\rangle\langle\, e^u_{\mathbf{z}} \,|\; = I. \tag{17}$$

Proof. We prove that $\| f \|^2_{\mathcal{H}_u} \equiv \langle\, f \,|\, f \,\rangle_{\mathcal{H}_u} = \|\hat{f}\|^2_{L^2}$. The inner product can be recovered by polarization. To begin with, assume that $\hat{f}$ is in the Schwartz space $\mathcal{S}(\mathbb{R}^s)$ of rapidly decreasing smooth test functions. Then

$$f_u(\mathbf{x} - i\mathbf{y}) = \Big(\exp\big[-u\mathbf{p}^2/2m + \mathbf{y}\cdot\mathbf{p}\big]\, \hat{f} \Big)^{\check{}}(\mathbf{x}), \tag{18}$$

hence by Plancherel's theorem,

$$\begin{aligned} &\int d^s\mathbf{x} \,|\, f_u(\mathbf{x} - i\mathbf{y}) \,|^{\,2} \\ &\qquad = (2\pi)^{-s} \int d^s\mathbf{p} \exp\big(-u\mathbf{p}^2/m + 2\mathbf{y}\cdot\mathbf{p}\big) \;|\, \hat{f}(\mathbf{p}) \,|^{\,2} \end{aligned} \tag{19}$$

and

$$\begin{aligned} &\int d\mu_u(\mathbf{z}) \,|\, f_u(\mathbf{z}) \,|^{\,2} \\ &\qquad = (m/\pi u)^{s/2} (2\pi)^{-s} \int d^s\mathbf{p} \exp\big[-u\mathbf{p}^2/m\big] \,|\, \hat{f}(\mathbf{p}) \,|^{\,2} \times \\ &\qquad\qquad \int d^s\mathbf{y} \exp\big(-m\mathbf{y}^2/u + 2\mathbf{y}\cdot\mathbf{p}\big) \\ &\qquad = (2\pi)^{-s} \int d^s\mathbf{p} \,|\, \hat{f}(\mathbf{p}) \,|^{\,2} = \|\, \hat{f} \,\|^2, \end{aligned} \tag{20}$$

where exchanging the order of integration was justified since the integrals are absolutely convergent. This proves (b), hence also (a), for $f \in \mathcal{S}(\mathbb{R}^s)$. Since the latter space is dense in $L^2(\mathbb{R}^s)$, the proof extends to $f \in L^2(\mathbb{R}^s)$ by continuity. (c) follows by noting that

$$\begin{aligned} \langle\, \hat{f} \,|\, \hat{g} \,\rangle_{L^2} &= \langle\, f \,|\, g \,\rangle_{\mathcal{H}_u} \\ &= \int d\mu_u(\mathbf{z}) \overline{f_u(\mathbf{z})} g_u(\mathbf{z}) \\ &= \int d\mu_u(\mathbf{z}) \langle\, \hat{f} \,|\, e^u_{\mathbf{z}} \,\rangle \langle\, e^u_{\mathbf{z}} \,|\, \hat{g} \,\rangle \end{aligned} \tag{21}$$

and dropping $\langle\, \hat{f} \,|$ and $|\, \hat{g} \,\rangle$. ∎

Using the map e^{-uH}, we can transfer any structure from $L^2(\mathbb{R}^s)$ to $\mathcal{H}_u$. In particular, time evolution is given by

$$\begin{aligned} f_u(\mathbf{z}, t) &= \left(e^{-uH}\left(e^{-itH} f\right)\right)(\mathbf{z}) \\ &= (2\pi)^{-s} \int d^s\mathbf{p}\, \exp\left[-i\tau\mathbf{p}^2/2m + i\mathbf{z}\cdot\mathbf{p}\right] \hat{f}(\mathbf{p}) \\ &\equiv \langle\, e_{\mathbf{z},\tau} \,|\, \hat{f} \,\rangle, \end{aligned} \tag{22}$$

where $\tau = t - iu$ and the wave packets

$$e_{\mathbf{z},\tau}(\mathbf{p}) = (2\pi)^{-s} \exp[-\bar{\tau}\mathbf{p}^2/2m - i\mathbf{p}\cdot\bar{\mathbf{z}}] \tag{23}$$

are obtained from the $e^u_{\mathbf{z}}$'s by evolving in real time t. They cannot be of minimal uncertainty since the free-particle Schrödinger equation is neccessarily dissipative. Instead, they give the following expectations and uncertainties:

$$\begin{aligned} \langle\, P_k \,\rangle &= (m/u) y_k \\ \langle\, X_k \,\rangle &= x_k - (t/u) y_k = x_k - (t/m) \langle\, P_k \,\rangle \\ \Delta_{P_k} &= \sqrt{m/2u} \\ \Delta_{X_k} &= \sqrt{\frac{u}{2m}\left(1 + \frac{t^2}{u^2}\right)}. \end{aligned} \tag{24}$$

Since

$$e_{\mathbf{z},\tau} = e^{itH} e^u_{\mathbf{z}}, \tag{25}$$

it follows that

$$\begin{aligned} \|f\|_\tau^2 &\equiv \int_{\mathbb{C}^s} d\mu_u(\mathbf{z}) \,|\, f(\mathbf{z},\tau) \,|\,^2 \\ &= \|e^{-it\mathbf{p}^2/2m}\hat{f}\|^2 = \|\hat{f}\|^2, \end{aligned} \tag{26}$$

thus we have a frame $\{\, e_{\mathbf{z},\tau} \,|\, \mathbf{z} \in \mathbb{C}^s \,\}$ at each complex "instant" $\tau = t - iu$, with the corresponding resolution of unity

$$\int_{\mathbb{C}^s} d\mu_u(\mathbf{z}) \,|\, e_{\mathbf{z},\tau} \,\rangle\langle\, e_{\mathbf{z},\tau} \,|\, = I. \tag{27}$$

The space $L^2(\mathbb{R}^s)$ carries a representation of the quantum mechanical Galilean group $\mathcal{G}_2$. Since the $e_{\mathbf{z},\tau}$'s were obtained from the dynamics associated with this group, they transform naturally under its action. A typical element of $\mathcal{G}_2$ has the form $g = (R, \mathbf{v}, \mathbf{x}_0, t_0, \theta)$, where R is a rotation, $\mathbf{v}$ is a boost, $\mathbf{x}_0$ is a spatial translation, t_0 is a time–translation and θ is the "phase" parameter associated with the central element $M = m/\hbar \equiv m$ in our representation (see section 4.2). g acts on the complex space–time domain $\mathcal{D}$ by sending the point $(\mathbf{z}, \tau)$ to $(\tau', \mathbf{z}')$, where

$$\begin{aligned} \mathbf{x}' &= R\mathbf{x} + t\mathbf{v} + \mathbf{x}_0 \\ \mathbf{y}' &= R\mathbf{y} + u\mathbf{v} \\ t' &= t + t_0 \\ u' &= u. \end{aligned} \tag{28}$$

The parameter θ has no effect on space–time; it only acts on wave functions by multiplying them by a phase factor. The representation of $\mathcal{G}_2$ is defined by

$$\big(U_g f\big)(\mathbf{z},\tau) = e^{-im\theta} f\big(g^{-1}(\mathbf{z},\tau)\big). \tag{29}$$

Thus we have

$$U_g \, e_{\mathbf{z},\tau} = e^{im\theta} e_{g(\mathbf{z},\tau)}, \tag{30}$$

and the $e_{\mathbf{z},\tau}$'s are "projectively covariant" under the action of $\mathcal{G}_2$; if we define $e_{\mathbf{z},\tau,\phi} \equiv e^{-im\phi} e_{\mathbf{z},\tau}$, then this expanded set is invariant under the action of $\mathcal{G}_2$, with $\phi' = \phi - \theta$. Since $e_{\mathbf{z},\tau}$ and $e_{\mathbf{z},\tau,\phi}$ represent the same physical state, we won't be fussy and just work with the

$e_{\mathbf{z},\tau}$'s. Anyway, this anomaly will disappear when we construct the corresponding relativistic coherent states.

The above representation of $\mathcal{G}_2$ on $L^2(\mathbb{R}^s)$ can be transfered to $\mathcal{H}_u$ using the map e^{-uH}. This map therefore *intertwines* (see Gelfand, Graev and Vilenkin [1966]) the representations on $\mathcal{G}_2$ on $L^2(\mathbb{R}^s)$ with the one on $\mathcal{H}_u$.

We conclude with some general remarks.

1. Since the $e^u_{\mathbf{z}}$'s are spherical and therefore invariant under $SO(n)$ (which is, after all, why they describe spinless particles), they can be parametrized by the homogeneous space $\mathcal{W} = \mathcal{G}_1/SO(n)$ as long as we keep u fixed (u is a parameter associated with the Hamiltonian, which is a generator of $\mathcal{G}_2$ but not of $\mathcal{G}_1$). The action of $\mathcal{W}$ as a *subgroup* of $\mathcal{G}_1$ on the $e^u_{\mathbf{z}}$'s is preserved in passing to the homogeneous space, hence $\mathcal{W}$ acts to translate these vectors in phase space. This explains the similarity between the $e^u_{\mathbf{z}}$'s and the canonical coherent states. On the other hand, *dynamics* (in imaginary time) is responsible for the parameter u. If we write $\mathbf{k} \equiv (m/u)\mathbf{y}$, then

$$\begin{aligned} e^u_{\mathbf{z}}(\mathbf{p}) &= (2\pi)^{-s} \exp\left[\frac{u}{2m}\mathbf{k}^2 - \frac{u}{2m}(\mathbf{p}-\mathbf{k})^2 - i\mathbf{p}\cdot\mathbf{x}\right] \\ &\equiv \exp\bigl[u\mathbf{k}^2/2m\bigr] e^{-i\mathbf{p}\cdot\mathbf{x}}\, h\left(\frac{\mathbf{p}-\mathbf{k}}{\sqrt{2m/u}}\right). \end{aligned} \tag{31}$$

The measure $d\mu_u(\mathbf{z})$ is now

$$d\mu_u(\mathbf{x},\mathbf{k}) = (u/\pi m)^{s/2} \exp\bigl(-u\mathbf{k}^2/m\bigr)\, d^s\mathbf{x}\, d^s\mathbf{k}. \tag{32}$$

Hence the exponential factor $\exp[u\mathbf{k}^2/2m]$ in $e^u_{\mathbf{z}}$, when squared in the reconstruction formula, precisely cancels the Gaussian weight factor in $d\mu_u(\mathbf{x},\mathbf{k})$, leaving the measure

$$\bigl(u/\pi m\bigr)^{s/2}\, d^s\mathbf{x}\, d^s\mathbf{k} \tag{33}$$

in phase space. It follows from the above form of $e^u_{\mathbf{z}}$ that $2\Delta_{P_k} = \sqrt{2m/u}$ plays the role of a *scale factor* in momentum space (as used in the wavelet transforms of chapter 1), hence its reciprocal $\Delta_{X_k} = \sqrt{u/2m}$ acts as a scale factor in configuration space. Thus the Galilean coherent states combine the properties of rigid "windows" with those of wavelets, due to the fact that their analytic semigroup $\mathcal{G}_2^c$ includes both phase–space translations and scaling, the latter due to the

heat operator e^{-uH}. However, note that u is constant, though arbitrary, in the resolution of unity and the corresponding reconstruction formula. Since there is an abundance of "wavelets" due to translations in phase space, only a single scale is needed for reconstruction. (One could, of course, include a range of scales by integrating over u with a weight function, but this seems unnecessary.) In the treatment of relativistic particles, u becomes the time component of a four–vector $y = (u, \mathbf{y})$, hence will no longer be constant. This is because *relativistic windows* shrink in the direction of motion, due to Lorentz contractions, thus automatically adjusting to the analysis of high–frequency components of the spectrum.

2. Notice that $e^u_{\mathbf{z}}$ is essentially the heat operator e^{-uH} applied to the δ–function at x, then analytically continued to $\bar{\mathbf{z}} = \mathbf{x} + i\mathbf{y}$. The fact that all the $e^u_{\mathbf{z}}$'s have minimal uncertainties shows that the action of the heat semigroup $\{U(-iu)\}$ is such that while the position undergoes the normal diffusion, the momentum undergoes the opposite process of *refinement,* in just such a way that the product of the two variances remains constant. This is reflected in the fact that the operator e^{-uH}, whose inverse in $L^2(\mathbb{R}^s)$ is unbounded, becomes unitary when the functions in its range get analytically continued, and the reconstruction formula is just a way of inverting e^{-uH}. Hence no information is lost if one looks in phase space rather than configuration space! It seems to me that this way of "inverting" semigroups must be an example of a general process. If such a process exists, I am unaware of it. In our case, at least, it appears to be possible because of analyticity.

3. So far, it seems that coherent–state representations are intimately connected with groups and their representations. However, there is a reasonable chance that coherent–state representations similar to the above can be constructed for systems which, unlike free particles, do not possess a great deal of symmetry. Suppose we are given a system of $s/3$ particles in $\mathbb{R}^3$ which interact with one another and/or with an external source through a potential $V(\mathbf{x})$. We assume that $V(\mathbf{x})$ is time–independent, so the system is conservative. (This means that we do have one symmetry, namely under time translations. If, moreover, the potential depends only on the differences $\mathbf{x}_i - \mathbf{x}_j$ between individual particles, we also have symmetry with respect to translations of the center of mass of the entire system; but we do not make this assumption here.) This system is then described by a Schrödinger equation with the Hamiltonian operator $H = H_0 + V$, where H_0 is the free Hamiltonian and V is the operator of multiplication by $V(\mathbf{x})$. We

need to assume that this (unbounded) operator can be extended to a self–adjoint operator on $L^2(\mathbb{R}^s)$. How far can the above construction be carried in this case? The key to our method was the positivity of the free Hamiltonian $H_0 = \mathbf{P}^2/2m$. But a general Hamiltonian must at least satisfy the *stability condition:*

(S) The spectrum of H is bounded below.

If H fails to meet this condition, then the system it describes is unstable, and a small perturbation can make it cascade down, giving off an infinite amount of energy. For a stable system, the evolution group $U(t) = e^{-itH}$ can be analytically continued to an analytic semigroup $U(\tau) = e^{-i\tau H}$ in the lower–half complex time plane as in the free case. Depending on the strength of the potential, the functions $f_u = U(-iu)f$ may be continued to some subset of $\mathbb{C}^s$. Formally, this corresponds to defining

$$\begin{aligned} f(\mathbf{z},\tau) &= \left(e^{i\mathbf{z}\cdot\mathbf{P}}e^{-i\tau H}f\right)(\mathbf{0}), \quad \tau = t - iu \\ &= \left(e^{\mathbf{y}\cdot\mathbf{P}}e^{-i\tau H}f\right)(\mathbf{x}) \end{aligned} \tag{34}$$

for an initial function $f(\mathbf{x})$ in $L^2(\mathbb{R}^s)$. As mentioned, this expression is formal since the operator $e^{\mathbf{y}\cdot\mathbf{P}}$ is unbounded and $e^{-i\tau H}f$ may not be in its domain. But it can make sense operating on the range of $e^{-i\tau H}$, which coincides with the range of e^{-uH}, provided $\mathbf{y}$ is not too large. Let $\mathcal{Y}_u$ be the set of all $\mathbf{y}$'s for which $e^{\mathbf{y}\cdot\mathbf{P}}$ is defined on the range of e^{-uH} and, furthermore, the function $\exp\left[i\mathbf{z}\cdot\mathbf{P}\right]\exp\left[-i\tau H\right]f$ is sufficiently regular to be evaluated at the origin in $\mathbb{R}^s$, no matter which initial f was chosen in $L^2(\mathbb{R}^s)$. For many potentials, of course, $\mathcal{Y}_u$ will consist of the origin alone; in that case there are no coherent states. We assume that $\mathcal{Y}_u$ contains at least some open neighborhood of the origin. Intuitively, we may think of $\mathcal{Y}_u$ as the set of all imaginary positions which can be attained by the particle in an imaginary time–interval u, while moving in the potential V. In the free case, $\mathcal{Y}_u = \mathbb{R}^s$ and there is no restriction on $\mathbf{y}$ provided only that $u > 0$. This corresponds to the fact that there is no "speed limit" for free non–relativistic particles, hence a particle can get to any imaginary position in a given positive imaginary time. For relativistic free particles, $\mathcal{Y}_u$ is the open sphere of radius uc, where c is the speed of light. Returning to our system of interacting particles, define the associated *complex space–time domain*

$$\mathcal{Z}_H = \{\, (\mathbf{x} - i\mathbf{y}, t - iu) \in \mathbb{C}^{s+1} \mid u > 0,\ \mathbf{y} \in \mathcal{Y}_u \,\}. \tag{35}$$

This is the set of all complex space–time points which can be reached by the system in the presence of the potential $V(\mathbf{x})$, and it is the label space for our prospective coherent states. These are now defined as *evaluation maps* on the space of analytically continued solutions:

$$f(\mathbf{z},\tau) = \langle\, e^H_{\mathbf{z},\tau} \,|\, f \,\rangle, \tag{36}$$

the inner product being in $L^2(\mathbb{R}^s)$. Then from the above expression, again formally, we have the *dynamical coherent states*

$$e^H_{\mathbf{z},\tau} = e^{i\bar\tau H} e^{-i\bar{\mathbf{z}}\cdot\mathbf{P}} \delta_{\mathbf{0}} \tag{37}$$

for $(\mathbf{z},\tau)$ in $\mathcal{Z}_H$.

What is still missing, of course, is the measure $d\mu^H_u$. (Since the potential is t–independent, so will be the measure, if it exists.) Finding the measure promises to be equally difficult to finding the propagator for the dynamics. The latter is closely related to the reproducing kernel,

$$K_H(\mathbf{z},\tau;\,\bar{\mathbf{z}}',\bar\tau') = \langle\, e^H_{\mathbf{z},\tau} \,|\, e^H_{\mathbf{z}',\tau'} \,\rangle. \tag{38}$$

K_H depends on τ and $\bar\tau'$ only through the difference $\tau - \bar\tau'$, and is the analytic continuation of the propagator to the domain $\mathcal{Z}_H \times \overline{\mathcal{Z}_H}$. It is related to the measure through the reproducing property,

$$\begin{aligned}\int_{\sigma_\tau} d\mu^H_u(\mathbf{z})\, K_H(\mathbf{z}',\tau';\,\bar{\mathbf{z}},\bar\tau)\, K_H(\mathbf{z},\tau;\,\bar{\mathbf{z}}'',\bar\tau'') \\ = K_H(\mathbf{z}',\tau';\,\bar{\mathbf{z}}'',\bar\tau''),\end{aligned} \tag{39}$$

where the integration is carried out over a "phase space" σ_τ in $\mathcal{Z}_H$ with a fixed value of $\tau = t - iu$. A reasonable candidate for $d\mu^H_u$ (see section 4.4) is

$$\begin{aligned} d\mu^H_u(\mathbf{z}) &= C\,\|e^H_{\mathbf{z},\tau}\|^{-2}\, d^s\mathbf{x}\, d^s\mathbf{y} \\ &\equiv C\, e^{-\phi_u(\mathbf{z})}\, d^s\mathbf{x}\, d^s\mathbf{y}. \end{aligned} \tag{40}$$

Rather than finding the measure explicitly, a more likely possibility is that its existence can be proved by functional–analytic methods for some classes of potentials and approximation techniques may be used to estimate it or at least derive some of its properties. The theoretical possibility that such a measure exists raises the prospect of an

interesting analogy between the quantum mechanics of a *single* system and a *statistical ensemble* of corresponding *classical* systems at equilibrium with a heat reservoir. In the case of a free particle, if we set $\mathbf{k} = (m/u)\mathbf{y}$ as above (see remark 1) and define T by $u = 1/2kT$ where k is Boltzmann's constant, then it so happens that our measure $d\mu_u$ is identical to the Gibbs measure for a classical canonical ensemble (see Thirring [1980]) of $s/3$ free particles of mass m in $\mathbb{R}^3$, at equilibrium with a heat reservoir at absolute temperature T. Thus, integrating with $d\mu_u$ over phase space is very much like taking the classical thermodynamic average at equilibrium! It remains to be seen, of course, whether this is a mere coincidence or if it has a generalization to interacting systems.

There is also a connection between the expectation values of an operator A in the coherent states $e^H_{\mathbf{z},\tau}$ and its thermal average in the Gibbs state,

$$\langle\, A \,\rangle_\beta \equiv Z^{-1}\,\mathrm{Trace}\,\left(e^{-\beta H}\,A\right) = Z^{-1}\,\mathrm{Trace}\,\left(e^{-\beta H/2}\,A\,e^{-\beta H/2}\right), \tag{41}$$

where $Z \equiv \mathrm{Trace}\,\left(e^{-\beta H}\right)$. Namely, if we have the resolution of unity

$$\int_{\sigma_\tau} d\mu_u^H(\mathbf{z})\;|\,e^H_{\mathbf{z},\tau}\,\rangle\langle\,e^H_{\mathbf{z},\tau}\,|\;= I, \tag{42}$$

then

$$\begin{aligned}\langle\, A \,\rangle_\beta &= Z^{-1}\int_{\sigma_\tau} d\mu_u^H(\mathbf{z})\,\langle\, e^H_{\mathbf{z},\tau}\,|e^{-\beta H/2}\,Ae^{-\beta H/2}\,e^H_{\mathbf{z},\tau}\,\rangle \\ &= Z^{-1}\int_{\sigma_\tau} d\mu_u^H(\mathbf{z})\,\langle\, e^H_{\mathbf{z},\tau-i\beta/2}\,A\,e^H_{\mathbf{z},\tau-i\beta/2}\,\rangle \\ &\equiv Z^{-1}\int_{\sigma_\tau} d\mu_u^H(\mathbf{z})\,\tilde{A}(\mathbf{z},\tau-i\beta/2),\end{aligned} \tag{43}$$

where we have used the formula

$$\mathrm{Trace}\,A = \int_{\sigma_\tau} d\mu_u^H(\mathbf{z})\,\langle\, e^H_{\mathbf{z},\tau}\,|\,A\,e^H_{\mathbf{z},\tau}\,\rangle, \tag{44}$$

which follows easily from eq. (42). Thus taking the thermal average means shifting the imaginary part u by $\beta/2$ in the integral.

4.4. Relativistic Frames

We are at last ready to embark on the main theme of this book: A new synthesis of Relativity and quantum mechanics through the geometry of complex spacetime. The main tool for this synthesis will be the physically necessary condition that the energy operator of the total system be non–negative, also known in quantum field theory as the *spectral condition.* The (unique) relativistically covariant statement of this condition gives rise to a canonical complexification of spacetime which embodies in its geometry the structure of quantum mechanics as well as that of Special Relativity. The complex spacetime also has the structure of a classical phase space underlying the quantum system under consideration. Quantum physics is developed through the construction of frames labeled by the complex spacetime manifold, which thus forms a natural bridge between the classical and quantum aspects of the system. It is hoped that this marriage, once fully developed, will survive the transition from Special to General Relativity.

As mentioned at the beginning of this chapter, the Perelomov–type constructions of chapter 3 do not apply directly to the Poincaré group since its time evolution (dynamics) is non–trivial. Pending a generalization of these methods to dynamical groups, we merely use the ideas of chapter 3 for inspiration rather than substance. In fact, it may well be that a closer examination of the construction to be developed here may suggest such a generalization.

We begin with the most basic object of relativistic quantum mechanics, the Klein–Gordon equation, which describes a simple relativistic particle in the same way that the Schrödinger equation describes a non–relativistic particle. The spectral condition will enable us to analytically continue the solutions of this equation to complex spacetime, and the evaluation maps on the space of these analytic solutions will be bounded linear functionals, giving rise to a reproducing kernel as in section 1.4. Physically, the evaluation maps are optimal wave packets, or coherent states, and it is this interpretation which establishes the underlying complex manifold as an extension of classical phase space. The next step is to build *frames* of such coherent states. (Recall from section 1.4 that a frame determines a reproducing kernel, but not vice versa.)

The coherent states we are about to construct are covariant under the restricted Poincaré group, hence they represent *relativistic wave packets* . As we have seen, such a covariant family is closely related to a unitary irreducible representation of the appropriate group,

in this case $\mathcal{P}_0$. Such representations are called *elementary systems*, and correspond roughly to the classical notion of *particles*, though with a definite quantum flavor. (For example, physical considerations prohibit them from being localized at a point in space, as will be discussed later.) We will focus on representations corresponding to *massive particles*. (A phase–space formalism for massless particles would be of great interest, but to my knowledge, no satisfactory formulation exists as yet.) Such representations are characterized by two parameters, the mass $m > 0$ and the spin $j = 0, 1/2, 1, 3/2, \ldots$ of the corresponding particle. We will specialize to spinless particles ($j = 0$) for simplicity. The extension of our construction to particles with spin is not difficult and will be taken up later. Thus we are interested in the (unique, up to equivalence) representation of $\mathcal{P}_0$ with $m > 0$ and $j = 0$. A natural way to construct this representation is to consider the space of solutions of the Klein–Gordon equation

$$(\Box + m^2c^2)f(x) = 0, \tag{1}$$

where

$$\begin{aligned} \Box &= c^{-2}\frac{\partial^2}{\partial t^2} - \Delta \\ &= \partial^\mu \partial_\mu \end{aligned} \tag{2}$$

is the Del'Ambertian, or wave operator, Δ is the usual spatial Laplacian and $\partial_\mu = \partial/\partial x^\mu$. The function f is to be complex–valued (for spin j, f is valued in $\mathbb{C}^{2j+1}$). We set $c = 1$ except as needed for future reference. If we write $f(x)$ as a Fourier transform,

$$f(x) = (2\pi)^{-s-1}\int_{\mathbb{R}^{s+1}} d^{s+1}p\; e^{-ixp}\,\tilde{f}(p), \tag{3}$$

then the Klein–Gordon equation requires that $\tilde{f}(p)$ be a distribution supported on the *mass shell*

$$\Omega_m = \{p = (p_0, \mathbf{p}) \in \mathbb{R}^{s+1} \,|\, p^2 \equiv p_0^2 - \mathbf{p}^2 = m^2\}. \tag{4}$$

Ω_m is a two–sheeted hyperboloid,

$$\Omega_m = \Omega_m^+ \cup \Omega_m^-, \tag{5}$$

where

$$p_0 = \pm\sqrt{m^2 + \mathbf{p}^2} \equiv \pm\omega(\mathbf{p}) \text{ on } \Omega_m^{\pm}. \tag{6}$$

Taking

$$\tilde{f}(p) = 2\pi\,\delta(p^2 - m^2)\,a(p) \tag{7}$$

for some function $a(p)$ on Ω_m, and using

$$\begin{aligned}\delta(p^2 - m^2) &= \delta\left((p_0 - \omega)(p_0 + \omega)\right) \\ &= \frac{1}{2\omega}\left[\delta(p_0 - \omega) + \delta(p_0 + \omega)\right],\end{aligned} \tag{8}$$

we get

$$\begin{aligned}f(x) &= \int_{\Omega_m} d\tilde{p}\; e^{-ixp}\,a(p) \\ &= \int_{\Omega_m^+} d\tilde{p}\;\left[e^{-ixp}\,a(p) + e^{ixp}\,a(-p)\right],\end{aligned} \tag{9}$$

where

$$d\tilde{p} = (2\pi)^{-s}(2\omega)^{-1}d^s\mathbf{p} \tag{10}$$

is the unique (up to a constant factor) Lorentz–invariant measure on Ω_m. (The factor ω^{-1} corrects for Lorentz contraction in frames at momentum p.) For physical particles, we must require that the energy be positive, i.e. that $a(p) = 0$ on Ω_m^-. Hence the physical states are given as *positive–energy solutions*,

$$f(x) = \int_{\Omega_m^+} d\tilde{p}\; e^{-ixp}\,a(p). \tag{11}$$

The function $a(p)$ can now be related to the initial data by setting $x^0 \equiv t = 0$, which shows that

$$\begin{aligned}f_0(\mathbf{x}) \equiv f(\mathbf{x}, 0) &= \int_{\Omega_m^+} d\tilde{p}\; e^{i\mathbf{x}\cdot\mathbf{p}}a(p) \\ &= \left((2\omega)^{-1}a\right)^{\check{}}(\mathbf{x}),\end{aligned} \tag{12}$$

so

$$a(p) \equiv a(\omega, \mathbf{p}) = 2\omega\hat{f}_0(\mathbf{p}), \tag{13}$$

where $\hat{}$ denotes the spatial Fourier transform. In particular, $f(x)$ is determined by its values on the Cauchy surface $t = 0$. For general solutions of the Klein–Gordon equation, we would also need to specify $\partial f/\partial t$ on that surface, but restricting ourselves to positive–energy solutions means that $f(x)$ actually satisfies the first–order pseudo–differential *non–local* equation

$$i\frac{\partial f}{\partial t} = \sqrt{m^2 - \Delta}\, f(x) \tag{14}$$

(which implies the Klein–Gordon equation), hence only $f(\mathbf{x}, 0)$ is necessary to determine f. (We will see that when analytically continued to complex spacetime, positive–energy solutions have a local characterization.) The inner product on the space of positive–energy solutions is defined using the Poincaré–invariant norm in momentum space,

$$\| f \|^2 \equiv \int_{\Omega_m^+} d\tilde{p} \;\; | \, a(p) \, |^2. \tag{15}$$

We will refer to the Hilbert space

$$L^2_+(d\tilde{p}) \equiv \{ a \in L^2(d\tilde{p}) \, | \, a(p) = 0 \text{ on } \Omega_m^- \} \tag{16}$$

as the space of positive–energy solutions in the *momentum representation*. It carries a unitary irreducible representation of $\mathcal{P}_0$ defined as follows. The natural action of $\mathcal{P}_0$ on spacetime is

$$(b, \Lambda)x = \Lambda x + b, \tag{17}$$

where Λ is a resticted Lorentz transformation ($\Lambda \in \mathcal{L}_0$) and b is a spacetime translation. Since the Klein–Gordon equation is invariant under $\mathcal{P}_0$, the induced action on functions over spacetime transforms solutions to solutions. Since the positivity of the energy is also invariant under $\mathcal{P}_0$, the subspace of positive–energy solutions is also left invariant. $\mathcal{P}_0$ acts on solutions by

$$\left(U(b, \Lambda) f\right)(x) \equiv f\left(\Lambda^{-1}(x - b)\right). \tag{18}$$

The invariance of the inner product on $L^2_+(d\tilde{p})$ then implies that the induced action on that space (which we denote by the same operator) is

$$\left(U(b, \Lambda)\, a\right)(p) = e^{ibp}\, a\left(\Lambda^{-1} p\right). \tag{19}$$

The invariance of the measure $d\tilde{p}$ then shows that $U(b,\Lambda)$ is unitary, thus $(b,\Lambda) \mapsto U(b,\Lambda)$ is a unitary representation of $\mathcal{P}_0$. It can be shown that it is, furthermore, irreducible.

Neither of the "function" spaces $\{f(x)\}$ and $L^2_+(d\tilde{p})$ are reproducing–kernel Hilbert spaces, since the evaluation maps $f \mapsto f(x)$ and $a \mapsto a(p)$ are unbounded. To obtain a space with bounded evaluation maps, we proceed as in the last section. Due to the positivity of the energy, solutions can be continued analytically to the lower–half time plane:

$$f(\mathbf{x}, t - iu) = \int_{\Omega_m^+} d\tilde{p}\ \exp\left(-it\omega - u\omega + i\mathbf{x}\cdot\mathbf{p}\right)\ a(p), \qquad (20)$$

where $u > 0$. As in the non–relativistic case, the factor $\exp(-u\omega)$ decays rapidly as $|\,\mathbf{p}\,| \to \infty$, which permits an analytic continuation of the solution to complex *spatial* coordinates $\mathbf{z} = \mathbf{x} - i\mathbf{y}$. But since $\omega(\mathbf{p}) \equiv \sqrt{m^2 + \mathbf{p}^2}$ is no longer quadratic in $|\,\mathbf{p}\,|$, $\mathbf{y}$ cannot be arbitrarily large. Rather, we must require that the four–vector $(u, \mathbf{y})$ satisfy the condition

$$u\omega - \mathbf{y}\cdot\mathbf{p} > 0 \qquad \forall(\omega, \mathbf{p}) \in \Omega_m^+. \qquad (21)$$

In covariant notation, setting $y^0 \equiv u$, we must have

$$yp > 0 \qquad \forall p \in \Omega_m^+, \qquad (22)$$

so that the complex exponential $\exp\left[-i(x - iy)p\right]$ remains bounded as p varies over Ω_m^+. This implies that $yp > 0$ for all $p \in V_+$, where

$$V_+ \equiv \{p = (p_0, \mathbf{p}) \in \mathbb{R}^{s+1} \mid\ |\mathbf{p}| < p_0/c\} \qquad (23)$$

is the *open forward light cone* in momentum space. In general, we need to consider the closure of V_+, i.e. the cone

$$\overline{V}_+ = \{p \in \mathbb{R}^{s+1} \mid |\mathbf{p}| \le p_0/c\}, \qquad (24)$$

which contains the light cone $\{p^2 = 0 \mid p_0 > 0\}$ (corresponding to massless particles) and the point $\{p = 0\}$ (corresponding in quantum field theory to the vacuum state). The set of all y's with $yp > 0$ is called the *dual cone* of $\overline{V}_+$, i.e.,

$$V'_+ \equiv \{y \in \mathbb{R}^{s+1} \mid\ yp > 0\ \forall p \in \overline{V}_+\}. \qquad (25)$$

It is easily seen that y belongs to V'_+ if and only if $|\,\mathbf{y}\,| < cy^0$. Note that as $c \to \infty$, $\overline{V}_+$ contracts to the non–negative p_0–axis while V'_+ expands to the half–space $\{(u,\mathbf{y}) \,|\, u > 0,\ \mathbf{y} \in \mathbb{R}^s\}$ which we have encountered in the last section. V'_+ coincides with V_+ when $c = 1$, but it is important to distinguish between them since they "live" in different spaces (see section 1.1).

Thus for $y \in V'_+$, setting $z = x - iy$, we define

$$f(z) = \int_{\Omega_m^+} d\tilde{p}\; e^{-izp}\, a(p). \tag{26}$$

The integral converges absolutely for any $a \in L^2_+(d\tilde{p})$ and defines a function on the *forward tube*

$$\mathcal{T}_+ \equiv \{x - iy \in \mathbb{C}^{s+1} \,|\, y \in V'_+\}, \tag{27}$$

also known as the *future tube* and, in the mathematical literature, as the *tube over* V'_+. Differentiation with respect to z^μ under the integral sign leaves the integral absolutely convergent, hence the function $f(z)$ is *holomorphic,* or analytic, in $\mathcal{T}_+$. As $y \to 0$ in V'_+, $f(z) \to f(x)$ in the sense of $L^2_+(d\tilde{p})$. Thus $f(x)$ is a *boundary value* of $f(z)$. Clearly $f(z)$ is a solution of the Klein–Gordon equation in either of the variables z or x. Let $\mathcal{K}$ be the space of all such holomorphic solutions:

$$\mathcal{K} = \{f(z) \,|\, a \in L^2_+(d\tilde{p})\}. \tag{28}$$

Then the map $a(p) \mapsto f(z)$ is one–to–one from $L^2_+(d\tilde{p})$ onto $\mathcal{K}$. Hence we can make $\mathcal{K}$ into a Hilbert space by defining

$$\langle\, f_1 \,|\, f_2 \,\rangle_{\mathcal{K}} \equiv \langle\, a_1 \,|\, a_2 \,\rangle, \tag{29}$$

where the inner product on the right–hand side is understood to be that of $L^2_+(d\tilde{p})$. We now show that $\mathcal{K}$ is a reproducing–kernel Hilbert space. Its evaluation maps are given by

$$E_z(f) \equiv f(z) = \int_{\Omega_m^+} d\tilde{p}\; e^{-izp}\, a(p) \equiv \langle\, e_z \,|\, a \,\rangle, \tag{30}$$

where

$$e_z(p) = e^{i\bar{z}p}. \tag{31}$$

Lemma 4.2.

1. *For each $z \in \mathcal{T}_+$, e_z belongs to $L^2_+(d\tilde{p})$, with*

$$\|e_z\|^2 = (2\pi)^{-1} \left(\frac{mc}{4\pi\lambda}\right)^{\nu} K_\nu(2\lambda mc), \tag{32}$$

where $\nu = (s-1)/2$,

$$\lambda \equiv \sqrt{y^2} = \sqrt{c^2(y^0)^2 - \mathbf{y}^2} > 0 \tag{33}$$

and K_ν is a modified Bessel function (Abramowitz and Stegun [1964]; the speed of light has been inserted for future reference.)

2. *In particular, the evaluation maps on $\mathcal{K}$ are bounded, with*

$$|\, f(z)\,| \leq \|e_z\|\, \|f\|. \tag{34}$$

Proof. Set $c = 1$. (To recover c, replace m by mc in the end.) Then

$$\|e_z\|^2 = \int_{\Omega_m^+} d\tilde{p}\; e^{-2yp} \equiv G(y). \tag{35}$$

Since $G(y)$ is Lorentz–invariant and $y \in V'_+$, we can evaluate the integral in a Lorentz frame in which $y = (\lambda, \mathbf{0})$:

$$\begin{aligned} G(y) = G(\lambda, \mathbf{0}) &= \int_{\Omega_m^+} d\tilde{p}\; e^{-2\lambda\omega} \\ &= (2\pi)^{-s} \int \frac{d^s\mathbf{p}}{2\sqrt{m^2+\mathbf{p}^2}} \exp\left[-2\lambda\sqrt{m^2+\mathbf{p}^2}\right]. \end{aligned} \tag{36}$$

Set $\mathbf{p} = m\mathbf{q}$. Then

$$\begin{aligned} G(y) &= (2\pi)^{-s} m^{s-1} \int \frac{d^s\mathbf{q}}{2\sqrt{1+\mathbf{q}^2}} \exp\left[-2\lambda m\sqrt{1+\mathbf{q}^2}\right] \\ &= (2\pi)^{-s} m^{s-1} \frac{\pi^{s/2}}{\Gamma(s/2)} \int_0^\infty \frac{r^{s-1}dr}{\sqrt{1+r^2}} \exp\left[-2\lambda m\sqrt{1+r^2}\right] \\ &= \frac{m^{s-1}}{(4\pi)^{s/2}\Gamma(s/2)} \int_0^\infty dt\, \sinh^{s-1} t\; \exp\left[-2\lambda m \cosh t\right] \\ &= (2\pi)^{-1} \left(\frac{m}{4\pi\lambda}\right)^{\nu} K_\nu(2\lambda m). \blacksquare \end{aligned} \tag{37}$$

The reproducing kernel can be obtained by analytic continuation from $\|e_z\|^2$:

$$\begin{aligned} K(z', \bar{z}) &\equiv \langle\, e_{z'} \,|\, e_z \,\rangle \\ &= \int_{\Omega_m^+} d\tilde{p}\ \exp\left[-i(z' - \bar{z})p\right] \\ &= (2\pi)^{-1} \left(\frac{mc}{2\pi\eta}\right)^{\nu} K_\nu(\eta mc), \end{aligned} \tag{38}$$

where

$$\begin{aligned} \eta &\equiv \sqrt{-(z' - \bar{z})^2} \\ &= \left[(y' + y)^2 - (x' - x)^2 + 2i(y' + y)(x' - x)\right]^{1/2} \end{aligned} \tag{39}$$

is defined by analytic continuation from $z' = z$ (when $\eta = 2\lambda$) as follows: The square–root function is defined on the complex plane cut along the negative real axis. Since y and y' both belong to V_+', so does $y' + y$. Now the argument of the square root is real if and only if $(y' + y)(x' - x) = 0$, and this can happen only when $(x' - x)^2 \leq 0$. (Otherwise, either $x' - x$ or $x - x'$ belongs to V_+', hence $(y' + y)(x' - x)$ is positive or negative, respectively.) But then,

$$-(z' - \bar{z})^2 = (y' + y)^2 - (x' - x)^2 \geq (y' + y)^2 > 0. \tag{40}$$

Thus for $z', z \in \mathcal{T}_+$, the quantity $-(z' - z)^2$ cannot belong to the negative real axis, so η is well–defined.

The reproducing kernel is closely related to the analytically continued (Wightman) *2–point function* for the scalar quantum field of mass m (Streater and Wightman [1964]):

$$K(z', \bar{z}) = -i\Delta^+(z' - \bar{z}). \tag{41}$$

We will encounter this and other 2–point functions again in the next chapter, in connection with quantum field theory.

Note: We will be interested in the behavior of $\|\, e_z \,\|$ near the boundary of $\mathcal{T}_+$, i.e. when $\lambda \sim 0$. From the properties of K_ν it follows that

$$\|e_z\|^2 \sim \frac{\Gamma(\nu)}{(4\pi)^{\nu+1}}\, \lambda^{-2\nu} \text{ when } \lambda \sim 0. \tag{42}$$

In particular, the evaluation maps are no longer bounded when $\lambda = 0$.

$\mathcal{P}_0$ acts on $\mathcal{T}_+$ by a complex extension of its action on real spacetime, i.e.,

$$z' \equiv (b, \Lambda)\, z = \Lambda z + b. \tag{43}$$

This means that $x' = \Lambda x + b$ as before, and $y' = \Lambda y$. (This is consistent with the phase–space interpretation of $\mathcal{T}_+$ to be established below.) The induced action on $\mathcal{K}$ is therefore

$$\left(U(b, \Lambda) f\right)(z) = f\left(\Lambda^{-1}(z - b)\right). \tag{44}$$

This implies that the wave packets e_z transform covariantly under $\mathcal{P}_0$, i.e.

$$U(b, \Lambda)\, e_z = e_{\Lambda z + b}. \tag{45}$$

We have now established that the space $\mathcal{K}$ of holomorphic positive–energy solutions is a reproducing–kernel Hilbert space. Recall that picking out the positive–energy part of $f(x)$ was a non–local operation in *real* spacetime, involving the pseudodifferential operator $\sqrt{m^2 - \Delta}$. However, when extended to $\mathcal{T}_+$, such functions may be characterized locally, as simultaneaous solutions of the Klein–Gordon equation and the Cauchy–Riemann equations, since the negative–energy part of $f(x)$ does not have an analytic continuation to $\mathcal{T}_+$.

We now show that $\mathcal{T}_+$ may, in fact, be interpreted as an extended phase space for the underlying classical relativistic particles. Clearly, $x^\mu \equiv \Re\, z^\mu$ are the spacetime coordinates. Their relation to the expectation values of the relativistic (Newton–Wigner) position operators will be discussed below. We now wish to investigate the relation of the imaginary coordinates $y^\mu \equiv -\Im\, z^\mu$ to the energy–momentum vector. The bridge between the "classical" coordinates y^μ and the quantum–mechanical observables P_μ will be, as usual, the (future) coherent states e_z. Before getting involved in computations, let us take a closer look at these wave packets in order to get a qualitative picture. Since yp is Lorentz–invariant, it can be evaluated in a reference frame where $p = (mc^2, \mathbf{0})$. Thus

$$yp = y^0 mc^2 = \sqrt{\lambda^2 + \mathbf{y}^2}\, mc \geq \lambda mc, \tag{46}$$

with equality if and only if $\mathbf{y} = \mathbf{0}$, i.e. when y and p are *parallel*. This is a kind of *reverse* Schwarz inequality which holds in $V_+' \times \overline{V}_+$ under

the pairing provided by the Lorentzian scalar product. Thus for *fixed* $y \in V_+'$ and variable $p \in \Omega_m^+$, we have

$$| e_z(p) | \leq e^{-\lambda mc}, \tag{47}$$

the maximum being attained when and only when $p = (mc/\lambda)y \equiv p_y$. Hence we expect, roughly, that

$$\langle P_\mu \rangle \equiv \frac{\langle e_z \,|\, P_\mu e_z \rangle}{\langle e_z \,|\, e_z \rangle} \sim (mc/\lambda) y_\mu. \tag{48}$$

Therefore the vector y, while itself *not* an energy–momentum, acts as a *control vector* for the energy–momentum by *filtering out* p's which are "far" from p_y. The larger we take the parameter λ, the *finer* the filter. The expected energy–momentum can be computed exactly by noting that

$$\begin{aligned} \langle e_z \,|\, P_\mu e_z \rangle &= \int_{\Omega_m^+} d\tilde{p}\; p_\mu \, e^{-2yp} \\ &= -\frac{1}{2} \frac{\partial G(y)}{\partial y^\mu} \end{aligned} \tag{49}$$

where $G(y) = \|e_z\|^2$ as before. Since G depends on y only through the invariant quantity λ, we have

$$\begin{aligned} \langle P_\mu \rangle &= -\frac{1}{2} G^{-1} \frac{\partial \lambda}{\partial y^\mu} \frac{\partial G}{\partial \lambda} \\ &= \left[\frac{K_{\nu+1}(2\lambda mc)}{K_\nu(2\lambda mc)} \cdot \frac{mc}{\lambda} \right] y_\mu, \end{aligned} \tag{50}$$

where we have used the recurrence relation (Abramowitz and Stegun [1964])

$$-\frac{\partial}{\partial \lambda} \left(\lambda^{-\nu} K_\nu(2\lambda m) \right) = 2m\lambda^{-\nu} K_{\nu+1}(2\lambda m). \tag{51}$$

This verifies and corrects the above qualitative estimate. In view of the above relation, the hyperboloid

$$\Omega_\lambda^+ \equiv \{ y \in V_+' \,|\, y^2 = \lambda^2 \} \tag{52}$$

corresponds to the mass shell. Hence the submanifold

$$\mathcal{C}_\lambda = \{x - iy \in \mathcal{T}_+ \mid y^2 = \lambda^2\} \tag{53}$$

corresponds to the extended phase space $\mathcal{C}$ defined in section 4.2 which, we recall, was a homogeneous space of $\mathcal{P}_0$ that was interpreted as spacetime×velocity space. Let us define the *effective mass* m_λ of the particle on Ω_λ^+ by

$$(m_\lambda c)^2 \equiv \langle\, P_\mu \,\rangle \langle\, P^\mu \,\rangle \equiv \langle\, P \,\rangle^2. \tag{54}$$

Then

$$m_\lambda = m \cdot \frac{K_{\nu+1}(2\lambda mc)}{K_\nu(2\lambda mc)}, \tag{55}$$

and

$$\langle\, P_\mu \,\rangle = \frac{m_\lambda c}{\lambda} y_\mu. \tag{56}$$

We claim that $m_\lambda > m$, which can be seen as follows. For all $p, p' \in \Omega_m^+$ we have the "reverse Schwarz inequality" $pp' \geq m^2$, with equality if and only if $p = p'$. Hence

$$\begin{aligned} m_\lambda^2 &= \langle\, P \,\rangle^2 \\ &= G^{-2} \iint d\tilde{p}\, d\tilde{p}'\ pp' \exp(-2yp - 2yp') \\ &> m^2. \end{aligned} \tag{57}$$

This is a kind of "renormalization effect" due to the uncertainty, or fluctuation, of the energy–momentum in the state e_z. It appears to go in the "wrong direction" (i.e., $\langle\, P \,\rangle^2 > \langle\, P^2 \,\rangle$) for the same reason as does the inequality $pp' \geq m^2$, namely because of the Lorentz metric.

Thus $\langle\, P_\mu \,\rangle$ is proportional, by a y–dependent but $\mathcal{P}_0$–invariant factor, to y_μ. We may therefore consider the y_μ's as *homogeneous coodinates* for the direction of motion of the classical particle in (real) spacetime. Alternatively, the expectation of the velocity operator $\mathbf{P}/P_0$ can be shown to be $\mathbf{y}/y_0$. Thus of the $s+1$ coordinates y_μ, only s have a "classical" interpretation. It is important to understand that the parameter λ *has no relation to the mass;* it can be chosen to be an arbitrary positive number and has the physical dimensions of length. It is the relativistic counterpart of the parameter u encountered in connection with the non–relativistic coherent states, and its

significance will be studied later. At this point we simply note that λ *measures the invariant distance of* z *from the boundary of* $\mathcal{T}_+$. The larger λ, the more smeared out are the spatial features and the more refined are the features in momentum space. (Recall that the imaginary part u of the time played a similar role in the non–relativistic theory.)

Because the vector y is so fundamental to our approach, it deserves a name of its own. We will call it the *temper vector*. The name is motivated in part by the smoothing effect which y has on spacetime quantities, and also by the fact that y plays a role similar to that played by the inverse teperature $\beta = 1/kT$ in statistical mechanics; the latter controls the energy.

From the asymptotic properties of the K_ν's it follows that

$$\langle\, P_\mu \,\rangle \sim \begin{cases} (mc/\lambda)y_\mu, & \text{when } \lambda mc \to \infty; \\ (\nu/\lambda^2)y_\mu, & \text{when } \lambda mc \to 0. \end{cases} \tag{58}$$

This can be understood as follows: When $\lambda mc \to \infty$, e.g. $c \to \infty$ for fixed λm, we recover the non–relativistic results. For example, the expectations of the spatial momenta approach those in the non–relativistic coherent states:

$$\langle\, P_k \,\rangle \sim (mc/\lambda)y_k \equiv (m/u)y_k. \tag{59}$$

When $\lambda mc \to 0$, say $\lambda \to 0$ for fixed mc, then z approaches the boundary of $\mathcal{T}_+$. In that case, fluctuations take over and the expectations become independent of the mass m.

The relation of the spacetime parameters $x_\mu \equiv \Re z_\mu$ to the Newton–Wigner position operators is as follows. Since a fixed $f \in \mathcal{T}_+$ describes the entire history of the particle, the associated state does not change with time (i.e., the dynamics is already built in). This means that we are in the so–called *Heisenberg picture*, and time–behavior must be described by evolving the observables A:

$$A(t) = e^{itP_0}\, A(0)\, e^{-itP_0}. \tag{60}$$

The Newton–Wigner position operators are uniquely determined by a set of seemingly reasonable assumptions concerning the localizability of the particle (Newton and Wigner [1949], Wightman [1962]), and are given in the momentum representation *at time* $x_0 = 0$ by

$$\begin{aligned} X_k(0) &= \omega^{1/2}\, i\frac{\partial}{\partial p_k}\, \omega^{-1/2} \\ &= i\left(\frac{\partial}{\partial p_k} - \frac{p_k}{\omega^2}\right), \quad k = 1, 2, \ldots, s. \end{aligned} \tag{61}$$

Now choose $z = x - iy \in \mathcal{T}_+$ with $x_0 = 0$. Then

$$\begin{aligned} \langle\, e_z \,|\, X_k(0)\, e_z \,\rangle &= \int_{\Omega_m^+} d\tilde{p}\; \omega^{1/2}\, e^{-izp}\, i\frac{\partial}{\partial p_k}\left(\omega^{-1/2}\, e^{i\bar{z}p}\right) \\ &= \Re \int_{\Omega_m^+} d\tilde{p}\; \omega^{1/2}\, e^{-izp}\, i\frac{\partial}{\partial p_k}\left(\omega^{-1/2}\, e^{i\bar{z}p}\right) \\ &= x_k\, \|e_z\|^2, \end{aligned} \tag{62}$$

hence

$$\langle\, X_k(0) \,\rangle = x_k. \tag{63}$$

It must be noted, however, that the concept of position for relativistic particles is highly unsatisfactory. Not only are the position operators non–covariant (this would seem to require a time operator on equal footing with them, which would exclude the possibility of dynamics); but the very concept of localizability for such particles is fraught with difficulties. For example, eigenvectors of the Newton–Wigner position operators, known as "localized states," spread out from a single point at time x_0 to fill the entire universe an arbitrarily small time later, violating relativistic causality. (The same phenomenon in the non–relativistic theory presents no conceptual problem, since propagation velocities are unrestricted there.) Even much weaker notions of localization give rise to problems with causality (Hegerfeldt [1985]). In my opinion, it is best to admit that position is simply a non–relativistic concept, and in the relativistic theory events x should be regarded as mere *parameters* of the spacetime manifold. As such, our formalism extends them to $z = x - iy$, with the new parameters y playing the role of a control vector for the energy–momentum observables. Thus, in place of a set of pairs of canonically conjugate observables X_k, P_k, we have a set of observables P_μ and a dual set of complex *parameters* z^μ. The symmetry between position– and momentum operators in the non–relativistic theory was based on the Weyl–Heisenberg group,

and we have seen that this symmetry is "accidental," being broken in the transition to relativity.
Note: This further reinforces the idea discussed in section 4.2, that "group–theoretic quantization," i.e. the quantization of classical systems obtained by requiring states and observables to transform under unitary representations of the associated dynamical groups or central extensions thereof, is superior to "canonical quantization" (sending the classical phase–space observables to operators, with Poisson brackets going to commutators). #

Let us now consider the uncertainties in the energy– and momentum operators. More generally, we compute the correlation matrix

$$C_{\mu\nu} \equiv \langle\, P_\mu P_\nu \,\rangle - \langle\, P_\mu \,\rangle\langle\, P_\nu \,\rangle, \tag{64}$$

of which the uncertainties $\Delta^2_{P_\mu}$ are the diagonal elements. From

$$\langle\, e_z \,|\, P_\mu P_\nu e_z \,\rangle = \frac{1}{4}\frac{\partial^2 G}{\partial y^\mu \partial y^\nu} \tag{65}$$

it follows that

$$\begin{aligned} 4C_{\mu\nu} &= G^{-1}\frac{\partial^2 G}{\partial y^\mu \partial y^\nu} - G^{-2}\frac{\partial G}{\partial y^\mu}\,\frac{\partial G}{\partial y^\nu} \\ &= \frac{\partial^2 \ln G}{\partial y^\mu \partial y^\nu}. \end{aligned} \tag{66}$$

(Incidentally, this shows that the function $\ln G$ is of some interest in itself: its first partials are the expected momenta, while its second partials are the correlations.) A computation similar to that for $\langle\, P_\mu \,\rangle$ gives

$$C_{\mu\nu} = y_\mu y_\nu \frac{m^2}{\lambda^2}\left(\frac{K_{\nu+2}(2\lambda m)}{K_\nu(2\lambda m)} - \frac{m_\lambda^2}{m^2}\right) - g_{\mu\nu}\frac{m_\lambda}{2\lambda}. \tag{67}$$

Although this expression does not appear to be enlightening in any obvious way, the uncertainties Δ_{P_μ} can be estimated from it in various limits such as $\lambda m \to \infty$ and $\lambda m \to 0$.

The reproducing kernel by itself is of limited use. Although it makes it possible to establish the interpretation of $\mathcal{T}_+$ as an extended classical phase space, it does not provide us with a direct physical interpretation of the function values $f(z)$. The inner product in $\mathcal{K}$ is borrowed from $L^2_+(d\tilde{p})$, hence a probability interpretation exists, so

far, only in momentum space. In the standard formulation of Klein–Gordon theory, it is possible to define the inner product in configuration space, but the corresponding density turns out not to be positive, thus precluding a probabilistic interpretation. This is one of the well–known difficulties with the first–quantized Klein–Gordon theory, and is one of the reasons cited for the necessity to go to quantum field theory (second quantization). We will see that the phase–space approach *does* admit a covariant probabilistic picture of relativistic quantum mechanics, thus making the theory more complete even before second quantization. These topics will be discussed further in the next section and the next chapter. At this point we wish only to define an "autonomous" inner product on $\mathcal{K}$ as an integral over a "phase space" lying in $\mathcal{T}_+$. This will provide us with a normal frame of evaluation maps (chapter 1).

Recall that in the non–relativistic theory of the last section, the norm in the space $\mathcal{H}_u$ of holomorphic solutions was obtained by integrating $|\, f(\mathbf{z},\tau)\,|^2$ with respect to the Gaussian measure $d\mu_u(\mathbf{z})$ over the phase space $\tau \equiv t - iu = \text{constant}$. But now, for given $y^0 = u$, only those $\mathbf{y}$'s with $|\,\mathbf{y}\,| < cu$ belong to V_+'. That is, the particle can only travel a finite imaginary distance in a finite imaginary time. In view of the relation $\langle\, P_\mu \,\rangle \propto y_\mu$, the obvious candidate for a phase space is the set defined for given $t \in \mathbb{R}$ and $\lambda > 0$ by

$$\sigma \equiv \sigma_{t,\lambda} = \{x - iy \in \mathcal{T}_+ \mid\ x^0 = t,\ y^2 = \lambda^2\}. \tag{68}$$

Such sets are not covariant, but a covariant extension will be found in the next section. As for the measure, a Gaussian weight function (such as $\exp(-m\mathbf{y}^2/u)$, which occured in $d\mu_u(\mathbf{y})$) is no longer satisfactory since it cannot be covariant. It turns out that we do not need a weight function at all! This can be seen as follows: In the non–relativistic case, the shift to complex time was performed once and for all by the operator e^{-uH}. For fixed $u > 0$, the weight function served to correct for the non–unitary translation from the real point $\mathbf{x}$ in space to the complex point $\mathbf{z} = \mathbf{x} - i\mathbf{y}$. However, if we restrict ourselves to the subset σ, then a translation to imaginary space is necessarily accompanied by a translation in imaginary time. The analog of the above translation is $(t - i\lambda, \mathbf{x}) \mapsto (t - i\sqrt{\lambda^2 + \mathbf{y}^2}, \mathbf{x} - i\mathbf{y})$. The increase in y^0, it turns out, *precisely compensates* for the shift to complex space! This follows from the fact that the operator e^{-yP}, which affects the total shift to complex spacetime, is relativistically invariant, hence the point $\mathbf{y} = \mathbf{0}$ no longer plays a special role. We will show later that in the non–relativistic limit, we recover the weight

function naturally. Hence the Gaussian weight function associated with the Galilean coherent states (which, as we have seen, is closely related to that associated with the canonical coherent states) has its origin in the *geometry* of the relativistic phase space, i.e. in the curvature of the hyperboloid $\{y^2 = \lambda^2\}$.

For $\sigma \equiv \sigma_{t,\lambda}$ as above and $f \in \mathcal{K}$, define

$$\|f\|_\sigma^2 = \int_\sigma d\sigma \mid f(z) \mid^2, \tag{69}$$

where we parametrize σ by $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{2s}$ and the measure $d\sigma$ is given by

$$d\sigma(z) = A_\lambda^{-1} d^s\mathbf{x}\, d^s\mathbf{y} \tag{70}$$

with

$$\begin{aligned} A_\lambda &= \pi^{-1} \left(\frac{\pi\lambda}{mc}\right)^{\nu+1} K_{\nu+1}(2\lambda mc) \\ &= \left(\frac{2\pi\lambda}{mc}\right)^s \cdot \frac{m_\lambda}{m}\, G(\lambda). \end{aligned} \tag{71}$$

Then we have the folowing result.

Theorem 4.3. *For all $f \in \mathcal{K}$,*

$$\|f\|_\sigma = \|f\|_{\mathcal{K}}. \tag{72}$$

Proof. Assume, to begin with, that $a(\mathbf{p}) \equiv a(\omega, \mathbf{p})$ belongs to the Schwartz space $\mathcal{S}(\mathbb{R}^s)$ of rapidly decreasing test functions. Then

$$\begin{aligned} f(z) &= (2\pi)^{-s} \int d^s\mathbf{p}\, (2\omega)^{-1} \exp(-ixp - yp)\, a(\mathbf{p}) \\ &= \left[(2\omega)^{-1} \exp(-it\omega - yp)\, a\right]\check{}(\mathbf{x}). \end{aligned} \tag{73}$$

Hence, by Plancherel's theorem,

$$\int_{x^0=t} d^s\mathbf{x} \mid f(x - iy) \mid^2 = (2\pi)^{-s} \int_{\mathbb{R}^s} d^s\mathbf{p}\, (4\omega^2)^{-1} e^{-2yp} \mid a(p) \mid^2. \tag{74}$$

Exchanging the order of integration in the integral representing $\|f\|_\sigma^2$, we obtain

$$\|f\|_\sigma^2 = A_\lambda^{-1}(2\pi)^{-s}\int d^s\mathbf{p}\,(4\omega^2)^{-1}\,|\,a(p)\,|^{\,2}\int d^s\mathbf{y}e^{-2yp}. \tag{75}$$

We now evaluate

$$J(p) \equiv \int d^s\mathbf{y}\,e^{-2yp} \tag{76}$$

as follows: Consider all $s+1$ components of p as independent and *define* $m(p) \equiv \sqrt{p^2}$. From the integral computed earlier, i.e.

$$\begin{aligned} G(y) &\equiv (2\pi)^{-s}\int d^s\mathbf{p}\,(2\omega)^{-1}e^{-2yp} \\ &= (2\pi)^{-1}\left(\frac{m}{4\pi\lambda}\right)^{\nu}K_\nu(2\lambda m), \end{aligned} \tag{77}$$

we obtain by exchanging p and y (as well as m and λ):

$$\int d^s\mathbf{y}\,(2y_0)^{-1}e^{-2yp} = \left(\frac{\pi\lambda}{m}\right)^{\nu}K_\nu(2\lambda m). \tag{78}$$

Taking the partial derivative with respect to p^0 on both sides gives

$$\begin{aligned} J(p) &= -\frac{\partial}{\partial p^0}\left[\left(\frac{\pi\lambda}{m}\right)^{\nu}K_\nu(2\lambda m)\right] \\ &= -(2\pi\lambda^2)^\nu\frac{\partial}{\partial p^0}\left[\xi^{-\nu}K_\nu(\xi)\right] \\ &= -(2\pi\lambda^2)^\nu\frac{\partial\xi}{\partial p^0}\frac{\partial}{\partial\xi}\left[\xi^{-\nu}K_\nu(\xi)\right], \end{aligned} \tag{79}$$

where $\xi(p) \equiv 2\lambda m(p)$. Using again the recurrence relation for the K_ν's, we get

$$J(p) = 2p_0\,A_\lambda. \tag{80}$$

Thus

$$\|f\|_\sigma^2 = (2\pi)^{-s}\int d^s\mathbf{p}\,(2\omega)^{-1}\,|\,a(p)\,|^{\,2} = \|f\|_{\mathcal{K}}^2 \tag{81}$$

for $a(\mathbf{p}) \in \mathcal{S}(\mathbb{R}^s)$. By continuity, this extends to all $a \in L^2_+(d\tilde{p})$ since $\mathcal{S}(\mathbb{R}^s)$ is dense in $L^2_+(d\tilde{p})$. ∎

If we define

$$\langle\, f_1 \,|\, f_2 \,\rangle_\sigma \equiv \int_\sigma d\sigma(z)\, \overline{f_1(z)} f_2(z), \tag{82}$$

then by polarization we have the following immediate consequence of the theorem.

Corollary 4.4. *For all $f_1, f_2 \in \mathcal{K}$,*

$$\langle\, f_1 \,|\, f_2 \,\rangle_\sigma = \langle\, f_1 \,|\, f_2 \,\rangle_{\mathcal{K}}. \tag{83}$$

In particular, $\langle \cdot \,|\, \cdot \rangle_\sigma$ defines a $\mathcal{P}_0$–invariant inner product on $\mathcal{K}$, and we have the resolution of unity

$$\int_\sigma d\sigma_\lambda(z)\; |\, e_z \,\rangle\langle\, e_z \,| \;= I, \tag{84}$$

making the vectors e_z with $z \in \sigma$ a normal frame. ∎

Note: The above results extend to the case $\lambda = 0$ by continuity. The normalization constant in the measure $d\sigma$ is then given by

$$A_0 \equiv \lim_{\lambda \downarrow 0} A_\lambda = \frac{\pi^\nu \Gamma(\nu+1)}{2(mc)^{s+1}}. \tag{85}$$

The same formula applies for fixed $\lambda > 0$ and $m \sim 0$, and shows that A_λ becomes unbounded as $m \to 0$. #

The vectors e_z belong to $L^2_+(d\tilde{p})$, but correspond to vectors $\tilde{e}_z$ in $\mathcal{K}$ defined by

$$\tilde{e}_z(z') = \langle\, e_{z'} \,|\, e_z \,\rangle = K(z' - \bar{z}). \tag{86}$$

The norm $\|f\|^2_\sigma \equiv \langle\, f \,|\, f \,\rangle_\sigma$ on $\mathcal{K}$ provides us with an interpretation of $|\, f(z) \,|^{\,2}$ as a *probability density* with respect to the measure $d\sigma_\lambda$ on the phase space σ. Within this interpretation, the wave packets e_z have the following optimality property: For fixed $z \in \mathcal{T}_+$ let

$$\hat{e}_z(z') = \frac{\langle\, e_{z'} \,|\, e_z \,\rangle}{\|e_z\|}. \tag{87}$$

Proposition 4.5. *Up to a constant phase factor, the function $\hat{e}_z$ is the unique solution to the following variational problem: Find $f \in \mathcal{K}$ such that $\|f\| = 1$ and $|\, f(z) \,|$ is a maximum.*

Proof. This follows at once from the Schwarz inequality and theorem 4.3, since by eq. (26),

$$\begin{aligned} |\, f(z)\,| &= |\,\langle\, e_z \,|\, a\,\rangle\,| = |\,\langle\, \tilde{e}_z \,|\, f\,\rangle\,| \\ &\leq \|\tilde{e}_z\|\,\|f\| = \|e_z\|\,\|f\|, \end{aligned} \tag{88}$$

with equality if and only if f is a constant multiple of $\tilde{e}_z$. ∎
According to our probability interpretation of $|\, f(z)\,|^2$, this means that the normalized wave packet $\hat{e}_z$ maximizes the probability of finding the particle at z.
Note: Unlike the non–relativistic coherent states of the last section, the e_z's do not have minimum uncertainty products. In fact, since the uncertainty product is not a Lorentz–invariant notion, it is *a priori* impossible to have relativistic coherent states with minimum uncertainty products. The above optimality, which *is* invariant, may be regarded as a reasonable substitute. Actually, there are better ways to measure uncertainty than the standard one used in quantum mechanics, which is just the variance. From a statistical point of view, the variance is just the second moment of the probability distribution. Perhaps the *best* definition of uncertainty, which includes all moments, is in terms of *entropy* (Bialynicki–Birula and Mycielski [1975], Zakai [1960]). Being necessarily non–linear, however, makes this definition less tractable.

4.5. Geometry and Probability

The formalism of the last section was based on the phase space $\sigma_{t,\lambda} \equiv \sigma$ and the measure $d\sigma$, neither of which is invariant under the action of $\mathcal{P}_0$ on $\mathcal{T}_+$. Yet, the resulting inner product $\langle\, \cdot \,|\, \cdot\,\rangle_\sigma$ is clearly invariant. It is therefore reasonable to expect that σ and $d\sigma$ merely represent one choice out of many. Our purpose here is to construct a large natural class of such phase spaces and associated measures to which our previous results can be extended. This class will include σ and will be invariant under $\mathcal{P}_0$. In this way our formalism is freed from its dependence on σ and becomes manifestly covariant. As a byproduct, we find that positive–energy solutions of the Klein–Gordon equation give rise to a *conserved probability current*, so the probabilistic interpretation becomes entirely compatible with the spacetime geometry. As is well–known, no such compatibility is possible in the usual approach to Klein–Gordon theory.

We begin by regarding $\mathcal{T}_+$ as an extended phase space (symplectic manifold) on which $\mathcal{P}_0$ acts by canonical transformations. Candidates for phase space are $2s$–dimensional symplectic submanifolds $\sigma \subset \mathcal{T}_+$, and $\mathcal{P}_0$ maps different σ's into one another by canonical transformations. A submanifold of the "product" form $\sigma = S - i\Omega_\lambda^+$, where S (interpreted as a generalized configuration space) is an s–submanifold of (real) spacetime $\mathbb{R}^{s+1}$, turns out to be symplectic if and only if S is given by $x^0 = t\,(\mathbf{x})$ with $|\nabla t| \leq 1$, that is, if and only if S *is nowhere timelike*. (This is slightly larger than the class of all *spacelike* configuration spaces admissible in the standard theory.) The original $\sigma_{t,\lambda}$ corresponds to $t\,(\mathbf{x}) \equiv$constant. The results of the last section are extended to all such phase spaces of product form.

The action of $\mathcal{P}_0$ on $\mathcal{T}_+$ is not transitive but leaves each of the $(2s+1)$–dimensional submanifolds

$$\mathcal{T}_\lambda^+ \equiv \{x - iy \in \mathcal{T}_+ \mid y^2 = \lambda^2\} \tag{1}$$

invariant. Each $\mathcal{T}_\lambda^+$ is a homogeneous space of $\mathcal{P}_0$, with isotropy subgroup $SO(s)$, hence

$$\mathcal{T}_\lambda^+ \approx \mathcal{P}_0/SO(s). \tag{2}$$

Thus $\mathcal{T}_\lambda^+$ corresponds to the homogeneous space $\mathcal{C}$ of section 4.2 (where we had specialized to $s = 3$). In view of the considerations in sections 4.2 and 4.4, each $\mathcal{T}_\lambda^+$ can be interpreted as the product of spacetime with "momentum space". Phase spaces σ will be obtained by taking slices to eliminate the time variable.

On the other hand, we also need a covariantly assigned measure for each σ. The most natural way this can be accomplished is to begin with a single $\mathcal{P}_0$–*invariant symplectic form* on $\mathcal{T}_+$ and require that its restriction to each σ be symplectic. This will make each σ a symplectic manifold (which, in any case, it must be to be interpreted as a classical phase space) and thus provide it with a canonical (Liouville) measure. Thus we look for the most general 2–form α on $\mathcal{T}_+$ such that

(a) α is closed, i.e., $d\alpha = 0$;

(b) α is non–degenerate, i.e., the $(2s+2)$–form $\alpha^{s+1} \equiv \alpha \wedge \alpha \wedge \cdots \wedge \alpha$ vanishes nowhere;

(c) for every $g = (a, \Lambda) \in \mathcal{P}_0$, $g^*\alpha = \alpha$, where $g^*\alpha$ denotes the pull–back of α under g (see Abraham and Marsden [1978]). Since every $\mathcal{P}_0$–invariant function on $\mathcal{T}_+$ depends on z only through y^2, the most general invariant 2–form is given by

$$\alpha = \phi(y^2)\, dy_\mu \wedge dx^\mu + \psi(y^2)\, y_\mu y_\nu dy^\mu \wedge dx^\nu. \tag{3}$$

Now the restriction (pullback) of the second term to $\mathcal{T}_\lambda^+$ vanishes, since it contains the factor $y_\mu dy^\mu = d(y^2)/2$. Furthermore, the coefficient $\phi(y^2)$ of the first term is constant on $\mathcal{T}_\lambda^+$. Hence we may confine our attention to the form

$$\alpha = dy_\mu \wedge dx^\mu \tag{4}$$

without any essential loss of generality. This form is symplectic as well as invariant, hence it fulfills all of the above conditions. $\mathcal{T}_+$, together with α, is a symplectic manifold, and invariance means that each $g \in \mathcal{P}_0$ maps $\mathcal{T}_+$ into itself by a canonical transformation.

A general $2s$–dimensional submanifold σ of $\mathcal{T}_+$ will be a potential phase space only if the restriction, or pullback, of α to σ is a symplectic form. We denote this restriction by α_σ. Let σ be given by

$$\sigma = \{z \in \mathcal{T}_+ \mid s(z) = h(z) = 0\}, \tag{5}$$

where $s(z)$ and $h(z)$ are two real–valued, C^∞ (or at least C^1) functions on $\mathcal{T}_+$ such that $ds \wedge dh \neq 0$ on σ. For example, $\sigma_{t,\lambda}$ can be obtained from $s(z) = x^0 - t$ and $h(z) = y^2 - \lambda^2$. The pullback α_σ depends only on the submanifold σ, not on the particular choice of s and h.

Proposition 4.6. *The form α_σ is symplectic if and only if the Poisson bracket*

$$\{s, h\} \equiv \frac{\partial s}{\partial x^\mu}\frac{\partial h}{\partial y_\mu} - \frac{\partial h}{\partial x^\mu}\frac{\partial s}{\partial y_\mu} \neq 0 \tag{6}$$

everywhere on σ.

Proof. α_σ is closed since α is closed and $d(\alpha_\sigma) = (d\alpha)_\sigma$. Hence α_σ is symplectic if and only if it is non–degenerate, i.e. if and only if its s-th exterior power α^s vanishes nowhere on σ. Now $(\alpha_\sigma)^s$ equals the pullback of α^s to σ, and a straightforward computation gives

$$\alpha^s = s!\, \widehat{dy}^{\,\mu} \wedge \widehat{dx}_\mu, \tag{7}$$

where

$$\begin{aligned} \widehat{dy}^{\,\mu} &= (-1)^s\, dy_0 \wedge dy_1 \wedge \cdots \wedge dy_{\mu-1} \wedge dy_{\mu+1} \wedge \cdots \wedge dy_s \\ \widehat{dx}_\mu &= (-1)^s\, dx^s \wedge dx^{s-1} \wedge \cdots \wedge dx^{\mu+1} \wedge dx^{\mu-1} \wedge \cdots \wedge dx^0. \end{aligned} \tag{8}$$

($\widehat{dy}^{\,\mu}$ and $\widehat{dx}_\mu$ are essentially the Hodge duals (Warner [1971]) of dy_μ and dx^μ, respectively, with respect to the Minkowski metric.)
Let $\{u_1, \ldots, u_{2s}, v_1, v_2\}$ be a basis for the tangent space of $\mathcal{T}_+$ at $z \in \sigma$, with $\{u_1, \ldots, u_{2s}\}$ a basis for the tangent space σ_z of σ. Since ds and dh vanish on the vectors u_j,

$$\begin{aligned}(\alpha^s \wedge ds \wedge dh)(u_1, \ldots, u_{2s}, v_1, v_2) \\ &= \alpha^s(u_1, \ldots, u_{2s})\,(ds \wedge dh)(v_1, v_2) \\ &= \alpha^s_\sigma(u_1, \ldots, u_{2s})\,(ds \wedge dh)(v_1, v_2).\end{aligned} \tag{9}$$

By assumption, $(ds \wedge dh)(v_1, v_2) \neq 0$. Therefore α_σ is non–degenerate at z if and only if $\alpha^s \wedge ds \wedge dh \neq 0$ at z. But by eq. (7),

$$\alpha^s \wedge ds \wedge dh = s!\,\{s, h\}\,dy \wedge dx, \tag{10}$$

where

$$\begin{aligned}dy &= dy_0 \wedge \cdots \wedge dy_s \\ dx &= dx^s \wedge \cdots \wedge dx^0.\end{aligned} \tag{11}$$

Hence $\alpha^s_\sigma \neq 0$ at z if and only if $\{s, h\} \neq 0$ at z. ∎

Let us denote the family of all such symplectic submanifolds σ by Σ_0.

Proposition 4.7. *Let $\sigma \in \Sigma_0$ and $g \in \mathcal{P}_0$. Then $g\sigma \in \Sigma_0$ and the restriction $g\colon \sigma \to g\sigma$ is a canonical transformation from (σ, α_σ) to $(g\sigma, \alpha_{g\sigma})$.*

Proof. Let g^* denote the pullback map defined by g, taking forms on $g\sigma$ to forms on σ. Then the invariance of α implies that

$$g^*\alpha_{g\sigma} = \alpha_\sigma. \tag{12}$$

Hence $\alpha_{g\sigma}$ is non–degenerate. It is automatically closed since α is closed. Thus $g\sigma \in \Sigma_0$. To say that $g\colon \sigma \to g\sigma$ is a canonical transformation means precisely that α_σ and $\alpha_{g\sigma}$ are related as above. ∎

We will be interested mainly in the special case where $h(z) = y^2 - \lambda^2$ for some $\lambda > 0$ and $s(z)$ depends only on x. Then the s–dimensional manifold

$$S \equiv \{x \in \mathbb{R}^{s+1} \mid s(x) = 0\} \tag{13}$$

is a potential generalized configuration space, and σ has the "product" form

$$\sigma = S - i\Omega_\lambda^+ \equiv \{x - iy \in \mathcal{T}_+ \mid x \in S, y \in \Omega_\lambda^+\}. \tag{14}$$

The following result is physically significant in that it relates the pseudo–Euclidean geometry of spacetime and the symplectic geometry of classical phase space. It says that σ is a phase space if and only if S is a (generalized) configuration space.

Theorem 4.8. *Let $\sigma = S - i\Omega_\lambda^+$ be as above. Then (σ, α_σ) is symplectic if and only if*

$$\frac{\partial s}{\partial x^\mu}\,\frac{\partial s}{\partial x\mu} \geq 0, \tag{15}$$

that is, if and only if S is nowhere timelike.

Proof. On σ, we have

$$\{s, h\} = 2\frac{\partial s}{\partial x^\mu}\, y^\mu \neq 0, \tag{16}$$

and we may assume $\{s, h\} > 0$ without loss. For fixed $x \in S$, the above inequality must hold for all $y \in \Omega_\lambda^+$, hence for all $y \in V_+'$. This implies that the vector $\partial s/\partial x^\mu$ is in the dual $\overline{V}_+$ of V_+'. ∎

We denote the family of all σ's as above (i.e., with S nowhere timelike) by Σ. It is a subfamily of Σ_0 and is clearly invariant under $\mathcal{P}_0$. Note that Σ admits *lightlike* as well as *spacelike* configuration spaces, whereas the standard theory only allows spacelike ones.

We will now generalize the results of the last section to all $\sigma \in \Sigma$. The $2s$–form α_σ^s defines a positive measure on σ, once we choose an orientation (Warner [1971]) for σ. (This can be done, for example, by choosing an *ordered* set of vector fields on σ which span the tangent space at each point; the order of such a basis is a generalization of the idea of a "right–handed" coordinate system in three dimensions.) The appropriate measure generalizing $d\sigma$ of the last section is now defined as

$$d\sigma = (s!\, A_\lambda)^{-1}\alpha_\sigma^s. \tag{17}$$

$d\sigma$ is the restriction to σ of a $2s$–form defined on all of $\mathcal{T}_+$, which we also denote by $d\sigma$. (This is a mild abuse of notation; in particular, the "d" here must *not* be confused with exterior differentiation!) By eq. (7), we have

$$d\sigma = A_\lambda^{-1}\,\widehat{dy}^{\,\mu} \wedge \widehat{dx}_\mu. \tag{18}$$

We now derive a concrete expression for $d\sigma$. Since s obeys eq. (15) and $ds \neq 0$ on σ, we can solve $ds = 0$ (satisfied by the *restriction* of ds to σ) for dx^0 and substitute this into $\widehat{dx}_k$. This (and a similar procedure for y) gives

$$\begin{aligned} \widehat{dx}_\mu &= \left(\frac{\partial s}{\partial x^0}\right)^{-1} \frac{\partial s}{\partial x^\mu}\, \widehat{dx}_0 \\ \widehat{dy}^{\,\mu} &= \left(\frac{\partial h}{\partial y_0}\right)^{-1} \frac{\partial h}{\partial y_\mu}\, \widehat{dy}^{\,0} = (y^\mu/y^0)\, \widehat{dy}^{\,0} \end{aligned} \tag{19}$$

on σ. Hence

$$d\sigma = A_\lambda^{-1} \left(y^0 \frac{\partial s}{\partial x^0}\right)^{-1} \left(y^\mu \frac{\partial s}{\partial x^\mu}\right) \widehat{dy}^{\,0} \wedge \widehat{dx}_0. \tag{20}$$

We identify σ with $\mathbb{R}^{2s}$ by solving $s(x) = 0$ for $x^0 = t\,(\mathbf{x})$ and mapping

$$(t\,(\mathbf{x}) - i\sqrt{\lambda^2 + \mathbf{y}^2},\, \mathbf{x} - i\mathbf{y}) \mapsto (\mathbf{x}, \mathbf{y}). \tag{21}$$

We further identify $\widehat{dy}^{\,0} \wedge \widehat{dx}_0$ with the Lebesgue measure $d^s\mathbf{y}\, d^s\mathbf{x}$ on $\mathbb{R}^{2s}$ (this amounts to choosing a non–standard orientation of $\mathbb{R}^{2s}$). Thus we obtain an expression for $d\sigma$ as a measure on $\mathbb{R}^{2s}$. Now $s(x) = 0$ on σ implies that

$$\begin{aligned} 0 &= \frac{\partial}{\partial x^k}\, s\,(t\,(\mathbf{x}), \mathbf{x}) \\ &= \frac{\partial s}{\partial x^0}\, \frac{\partial t}{\partial x^k} + \frac{\partial s}{\partial x^k}, \end{aligned} \tag{22}$$

which can be substituted into the above expression to give

$$\begin{aligned} d\sigma &= A_\lambda^{-1} \left(1 - \frac{\partial t}{\partial x^k}\, \frac{y^k}{y^0}\right) d^s\mathbf{y}\, d^s\mathbf{x} \\ &= A_\lambda^{-1} \big(1 - \nabla t \cdot (\mathbf{y}/y_0)\big)\, d^s\mathbf{y}\, d^s\mathbf{x}. \end{aligned} \tag{23}$$

But eq. (15) implies that $|\,\nabla t\,(\mathbf{x})\,| \leq 1$, hence for $y \in V_+'$,

$$\big|\nabla t \cdot (\mathbf{y}/y_0)\,\big| < 1 \tag{24}$$

and $d\sigma$ is a positive measure as claimed. The above also shows that if $|\nabla t(\mathbf{x})| = 1$ for some $\mathbf{x}$, then $d\sigma$ becomes "asymptotically" degenerate as $|\mathbf{y}| \to \infty$ in the direction of $\nabla t(\mathbf{x})$. That is, if σ is *lightlike* at $(t(\mathbf{x}), \mathbf{x})$, then $d\sigma$ becomes small as the velocity $\mathbf{y}/y_0$ approaches the speed of light in the direction of $\nabla t(\mathbf{x})$. This means that functions in $L^2(d\sigma)$ (and, in particular, as we shall see, in $\mathcal{K}$) are allowed high velocities in the direction $\nabla t(\mathbf{x})$ at $(t(\mathbf{x}), \mathbf{x}) \in S$. This argument is an example of the kind of *microlocal analysis* which is possible in the phase–space formalism. (In the usual spacetime framework, one cannot say anything about the velocity distribution of a function at a given point in spacetime, since this would require taking the Fourier transform and hence losing the spatial information.)

For $\sigma \in \Sigma$, denote by $L^2(d\sigma)$ the Hilbert space of all complex–valued, measurable functions on σ with

$$\|f\|_\sigma^2 \equiv \int_\sigma d\sigma \, |f|^2 < \infty. \tag{25}$$

If f is a C^∞ function on $\mathcal{T}_+$, we restrict it to σ and define $\|f\|_\sigma$ as above. Our goal is to show that $\|f\|_\sigma = \|f\|_\mathcal{K}$ for every $f \in \mathcal{K}$. To do this, we first prove that each $f \in \mathcal{K}$ defines a conserved current in spacetime, which, by Stokes' theorem, makes it possible to deform the phase space $\sigma_{t,\lambda}$ of the last section to an arbitrary $\sigma = S - i\Omega_\lambda^+ \in \Sigma$ without changing the norm. For $f \in \mathcal{K}$, define

$$J^\mu(x) = A_\lambda^{-1} \int_{\Omega_\lambda^+} \widehat{dy}^{\,\mu} \, |f(x - iy)|^2, \tag{26}$$

where Ω_λ^+ has the orientation defined by $\widehat{dy}^{\,0}$, so that $J^0(x)$ is positive. Then

$$\|f\|_\sigma^2 = \int_S \widehat{dx}_\mu \, J^\mu(x), \tag{27}$$

where S has the orientation defined by $\widehat{dx}_0$. (The restriction of $\widehat{dx}_0$ to S does not vanish since $|\nabla t(\mathbf{x})| \leq 1$.)

Theorem 4.9. *Let $\hat{f}(\mathbf{p})$ be C^∞ with compact support. Then $J^\mu(x)$ is C^∞ and satisfies the continuity equation*

$$\frac{\partial J^\mu}{\partial x^\mu} = 0. \tag{28}$$

Proof. By eq. (19),

$$J^\mu(x) = A_\lambda^{-1} \int_{\Omega_\lambda^+} d\tilde{y}\ y^\mu \,|\, f(x - iy) \,|^2, \tag{29}$$

where $d\tilde{y} \equiv \widehat{dy}^{\,0}/y^0$. The function

$$F_x^\mu(y, p, q) \equiv y^\mu \exp\left[ix(p - q) - y(p + q)\right]\, \overline{\hat{f}(p)}\, \hat{f}(q) \tag{30}$$

is in $L^1(d\tilde{y} \times d\tilde{p} \times d\tilde{q})$, hence by Fubini's theorem,

$$\begin{aligned} J^\mu(x) &= A_\lambda^{-1} \int_{\Omega_\lambda^+} d\tilde{y} \int_{\Omega_m^+ \times \Omega_m^+} d\tilde{p}\, d\tilde{q}\, F_x^\mu(y, p, q) \\ &= A_\lambda^{-1} \int_{\Omega_m^+ \times \Omega_m^+} d\tilde{p}\, d\tilde{q} \exp\left[ix(p - q)\right]\, \overline{\hat{f}(p)}\, \hat{f}(q)\, H^\mu(p + q), \end{aligned} \tag{31}$$

where, setting $k \equiv p + q$, $\eta \equiv \sqrt{k^2}$ and using the recurrence relation for the K_ν's given by eq. (51) in section 4.4, we compute

$$\begin{aligned} H^\mu(k) &\equiv \int_{\Omega_\lambda^+} d\tilde{y}\ y^\mu\, e^{-yk} \\ &= -\frac{\partial}{\partial k_\mu} \int d\tilde{y}\, e^{-yk} \\ &= -\frac{\partial}{\partial k_\mu} \left[2 \left(\frac{2\pi\lambda}{\eta} \right)^\nu K_\nu(\lambda\eta) \right] \\ &= (k^\mu/\pi) \left(\frac{2\pi\lambda}{\eta} \right)^{\nu+1} K_{\nu+1}(\lambda\eta) \\ &\equiv k^\mu H(\eta). \end{aligned} \tag{32}$$

$H(\eta)$ is a bounded, continuous function of η for $\eta \geq 2m$, and

$$J^\mu(x) = A_\lambda^{-1} \int_{\Omega_m^+ \times \Omega_m^+} d\tilde{p}\, d\tilde{q} \exp\left[ix(p - q)\right]\, \overline{\hat{f}(p)}\, \hat{f}(q)\ (p^\mu + q^\mu)\ H(\eta). \tag{33}$$

Since $\hat{f}(\mathbf{p})$ has compact support, differentiation under the integral sign to any order in x gives an absolutely convergent integral, proving that J^μ is C^∞. Differentiation with respect to x^μ brings down the

factor $i(p^\mu - q^\mu)$ from the exponent, hence the continuity equation follows from $p^2 = q^2 = m^2$. ∎

Remark. The continuity equation also follows from a more intuitive, geometric argument. Let

$$B_\lambda^+ = \{y \in V_+' \mid y^0 > \sqrt{\lambda^2 + \mathbf{y}^2}\}, \tag{34}$$

oriented such that

$$\Omega_\lambda^+ = -\partial B_\lambda^+ \tag{35}$$

(the outward normal on ∂B_λ^+ points "down," whereas Ω_λ^+ is oriented "up"). Then by Stokes' theorem,

$$\begin{aligned} J^\mu(x) &= A_\lambda^{-1} \int_{\Omega_\lambda^+} \widehat{dy}^{\,\mu} \mid f(x - iy) \mid^2 \\ &= -A_\lambda^{-1} \int_{B_\lambda^+} d\left(\widehat{dy}^{\,\mu} \mid f(x - iy) \mid^2\right). \end{aligned} \tag{36}$$

Here, d represents exterior differentiation with respect to y, and since the s–form $\widehat{dy}^{\,\mu}$ contains all the dy_ν's except for dy_μ, we have

$$J^\mu(x) = -A_\lambda^{-1} \int_{B_\lambda^+} dy \, \frac{\partial}{\partial y_\mu} \mid f(x - iy) \mid^2, \tag{37}$$

where dy is Lebesgue measure on B_λ^+. To justify the use of Stokes' theorem, it must be shown that the contribution from $\mid \mathbf{y} \mid \to \infty$ to the first integral vanishes. This depends on the behavior of $f(z)$, which is why we have given the previous analytic proof using the Fourier transform. Then the continuity equation is obtained by differentiating under the integral sign (which must also be justified) and using

$$\frac{\partial^2 \mid f \mid^2}{\partial x^\mu \partial y_\mu} = 0, \tag{38}$$

which follows from the Klein–Gordon equation *combined with analyticity*, since

$$\begin{aligned}\frac{\partial^2}{\partial x^\mu \partial y_\mu} &= \left(\frac{\partial}{\partial \bar{z}^\mu} + \frac{\partial}{\partial z^\mu}\right) \cdot i \left(\frac{\partial}{\partial \bar{z}_\mu} - \frac{\partial}{\partial z_\mu}\right) \\ &= i\frac{\partial^2}{\partial \bar{z}^\mu \partial \bar{z}_\mu} - i\frac{\partial^2}{\partial z^\mu \partial z_\mu} \\ &\equiv i\left(\Box_{\bar{z}} - \Box_z\right).\end{aligned} \tag{39}$$

Incidentally, this shows that

$$\begin{aligned}j^\mu(z) &\equiv -\frac{\partial}{\partial y_\mu} \mid f(z) \mid^2 \\ &= i\left[\overline{f(z)}\,\partial_\mu f(z) - \overline{\partial_\mu f(z)} \cdot f(z)\right]\end{aligned} \tag{40}$$

is a "microlocal," spacetime–conserved probability current for each fixed $y \in V_+'$, so the scalar function $\mid f(z) \mid^2$ is a *potential* for the probability current. We shall see that this is a general trend in the holomorphic formalism: many vector and tensor quantities can be derived from scalar potentials.

Eqs. (37) and (40) also show that our probability current is a *regularized version* of the usual current associated with solutions in *real* spacetime. The latter (Itzykson and Zuber [1980]) is given by

$$J^\mu_{\text{usual}}(x) = i\left[\overline{f(x)}\,\partial_\mu f(x) - \overline{\partial_\mu f(x)} \cdot f(x)\right],$$

which leads to a conceptual problem since the time component, which should serve as a probability density, can become negative even for positive–energy solutions (Gerlach, Gomes and Petzold [1967], Barut and Malin [1968]). By contrast, eq. (36) shows that $J^0(x)$ is stricly non–negative. The tendency of quantities in complex spacetime to give regularizations of their counterparts in real spacetime is further discussed in chapter 5.

We can now prove the main result of this section.

Theorem 4.10. *Let $\sigma = S - i\Omega_\lambda^+ \in \Sigma$ and $f \in \mathcal{K}$. Then $\|f\|_\sigma = \|f\|_{\mathcal{K}}$.*

Proof. We will prove the theorem for $\hat{f}(\mathbf{p})$ in the space $\mathcal{D}(\mathbb{R}^s)$ of C^∞ functions with compact support, which implies it for arbitrary $\hat{f} \in L^2_+(d\tilde{p})$ by continuity. Let S be given by $x^0 = t\,(\mathbf{x})$, and for $R > 0$ let

$$\begin{aligned} D_R &= \{x \in \mathbb{R}^{s+1} \mid |\mathbf{x}| < R, x^0 \in [0, t(\mathbf{x})]\}, \\ E_R &= \{x \in \mathbb{R}^{s+1} \mid |\mathbf{x}| = R, x^0 \in [0, t(\mathbf{x})]\}, \\ S_{0R} &= \{x \in \mathbb{R}^{s+1} \mid |\mathbf{x}| < R, x^0 = 0\}, \\ S_R &= \{x \in \mathbb{R}^{s+1} \mid |\mathbf{x}| < R, x^0 = t(\mathbf{x})\}, \end{aligned} \tag{41}$$

where $[0, t(\mathbf{x})]$ means $[t(\mathbf{x}), 0]$ if $t(\mathbf{x}) < 0$. We orient S_{0R} and S_R by $\widehat{dx}_0$, E_R by the "outward normal"

$$\hat{\mathbf{r}} = R^{-1} \sum_{k=1}^{s} x^k \widehat{dx}_k, \tag{42}$$

and D_R so that $\partial D_R = S_R - S_{0R} + E_R$. Now let $\hat{f}(\mathbf{p}) \in \mathcal{D}(\mathbb{R}^s)$. Then $J^\mu(x)$ is C^∞, hence by Stokes' theorem,

$$\begin{aligned} \int_{S_R - S_{0R} + E_R} J^\mu(x)\, \widehat{dx}_\mu &= \int_{D_R} d\left(J^\mu\, \widehat{dx}_\mu\right) \\ &= (-1)^s \int_{D_R} dx\, \frac{\partial J^\mu}{\partial x^\mu} = 0. \end{aligned} \tag{43}$$

We will show that

$$\Delta(R) \equiv \int_{E_R} J^\mu \widehat{dx}_\mu \to 0 \text{ as } R \to \infty \tag{44}$$

(i.e., there is no leakage to $|\mathbf{x}| \to \infty$), which implies that

$$\begin{aligned} \|f\|_\sigma^2 &\equiv \lim_{R\to\infty} \int_{S_R} J^\mu\, \widehat{dx}_\mu \\ &= \lim_{R\to\infty} \int_{S_{0R}} J^\mu\, \widehat{dx}_\mu \\ &= \|f\|_{\sigma_{0\lambda}}^2 = \|f\|_{\mathcal{K}}^2 \end{aligned} \tag{45}$$

by theorem 1 of section 4.4. To prove that $\Delta(R) \to 0$, note that on E_R, $\widehat{dx}_0 = 0$ and

$$\widehat{dx}_k = x_k \frac{\widehat{dx}_1}{x_1} = x_k \frac{\widehat{dx}_2}{x_2} = \cdots = x_k \frac{\widehat{dx}_s}{x_s}, \tag{46}$$

each form being defined except on a set of measure zero; hence

$$\hat{\mathbf{r}} = R\,\frac{\widehat{dx}_1}{x_1}. \tag{47}$$

By eq. (29),

$$|J^k(x)| \le J^0(x), \tag{48}$$

hence

$$\begin{aligned} \big|\Delta(R)\big| &= \Big|\sum_{k=1}^{s}\int_{E_R} J^k\,\widehat{dx}_k\Big| \\ &= \Big|\sum_{k=1}^{s}\int_{E_R} J^k\,x_k\,\frac{\widehat{dx}_1}{x_1}\Big| \\ &\le s\int_{E_R} J^0\,R\,\frac{\widehat{dx}_1}{x_1} \\ &= s\int_{E_R} J^0\,\hat{\mathbf{r}} \equiv a(R). \end{aligned} \tag{49}$$

Now by eqs. (31) and (32),

$$J^0(x) = \int_{\mathbb{R}^{2s}} d^s\mathbf{p}\,d^s\mathbf{q}\;e^{ix(p-q)}\,\phi(\mathbf{p},\mathbf{q}), \tag{50}$$

where

$$\phi(\mathbf{p},\mathbf{q}) = \overline{\hat{f}(p)}\,\hat{f}(q)\,\psi(\mathbf{p},\mathbf{q}), \tag{51}$$

and $\psi \in C^\infty(\mathbb{R}^{2s})$. Hence $\phi \in \mathcal{D}(\mathbb{R}^{2s})$. Let

$$D = \hat{\mathbf{x}}\cdot\nabla_{\mathbf{p}}, \tag{52}$$

where $\hat{\mathbf{x}} = \mathbf{x}/R$, and observe that for $\mathbf{x} \in E_R$,

$$\begin{aligned} De^{ixp} &= -iR\left(1 - \frac{x^0}{R}\hat{\mathbf{x}}\cdot\mathbf{v}\right)e^{ixp} \\ &\equiv -iR\,\xi(x,p)\,e^{ixp}, \end{aligned} \tag{53}$$

where $\mathbf{v} = \mathbf{p}/p_0$. Since ϕ has compact support, there exists a constant $\alpha < 1$ such that $|\mathbf{v}| \le \alpha$ and $|\mathbf{v}'| \le \alpha$ for all $(\mathbf{p},\mathbf{p}')$ in the support

of ϕ. Furthermore, since $|\nabla t\,(x)| \leq 1$, given any $\epsilon > 0$ we have $|x^0| < R(1+\epsilon)$ for $\xi \in E_R$ for R sufficiently large; hence

$$|\xi(x,p)| \geq 1 - \alpha(1+\epsilon) \text{ for } x \in E_R \text{ and } \mathbf{p} \in \text{ supp } \phi. \tag{54}$$

Choose $0 < \epsilon < \alpha^{-1} - 1$, substitute

$$e^{ixp} = \frac{i}{R\,\xi(x,p)}\, D\, e^{ixp}, \qquad x \in E_R \tag{55}$$

into the expression for $J^0(x)$ and integrate by parts:

$$\begin{aligned} J^0(x) &= (iR)^{-1} \int_{\mathbb{R}^{2s}} d^s\mathbf{p} d^s\mathbf{q}\, e^{ix(p-q)}\, D\left(\frac{\phi(\mathbf{p},\mathbf{q})}{\xi(x,p)}\right) \\ &\equiv (iR)^{-1} \int_{\mathbb{R}^{2s}} d^s\mathbf{p} d^s\mathbf{q}\, e^{ix(p-q)}\, \phi'_x(\mathbf{p},\mathbf{q}). \end{aligned} \tag{56}$$

This procedure can be continued, giving (for $x \in E_R$)

$$J^0(x) = (iR)^{-n} \int_{\mathbb{R}^{2s}} d^s\mathbf{p} d^s\mathbf{q}\, e^{ix(p-q)}\, \phi^{(n)}_x(\mathbf{p},\mathbf{q}), \qquad n = 1, 2, \cdots, \tag{57}$$

where

$$\begin{aligned} \phi^{(n)}_x(\mathbf{p},\mathbf{q}) &= \left(D \circ \xi^{-1}\right)^n \phi(\mathbf{p},\mathbf{q}) \\ &= \left[\hat{\mathbf{x}} \cdot \nabla_{\mathbf{p}} \left(1 - \frac{x^0}{R}\, \hat{\mathbf{x}} \cdot \mathbf{v}\right)^{-1}\right]^n \phi(\mathbf{p},\mathbf{q}). \end{aligned} \tag{58}$$

Now $\left(D \circ \xi^{-1}\right)^n$ is a partial differential operator in $\mathbf{p}$ whose coefficients are polynomials in $D^k(\xi^{-1})$ with $k = 0, 1, \cdots, n$. We will show that for $x \in E_R$ with R sufficiently large, there are constants b_k such that

$$\left|D^k(\xi^{-1})\right| < b_k, \qquad k = 0, 1, \cdots, \tag{59}$$

which implies that

$$\|\phi^{(n)}_x\|_{L^1(\mathbb{R}^{2s})} < c_n, \qquad x \in E_R,\ n = 1, 2, \cdots \tag{60}$$

for some constants c_n, so that by eqs (49) and (57),

$$\begin{aligned} a(R) &= s \int_{E_R} J^0 \hat{\mathbf{r}} \\ &=\leq sR^{-n} \int_{E_R} \|\phi_x^{(n)}\|_{L^1(\mathbb{R}^{2s})}\, \hat{\mathbf{r}}(x) \\ &\leq sR^{-n}\, c_n \frac{2\pi^{s/2}}{\Gamma(s/2)}\, R^{s-1} \int_0^{R(1+\epsilon)} dx^0 \\ &= \frac{2s\pi^{s/2}}{\Gamma(s/2)}\, c_n R^{s-n}\, (1+\epsilon) \\ &\to 0 \text{ as } R \to \infty \end{aligned} \tag{61}$$

if we choose $n > s$. To prove eq. (59), note that it holds for $k = 0$ by eq. (54) and let $u = \hat{\mathbf{x}} \cdot \mathbf{v}$. Then

$$Du \equiv (\hat{\mathbf{x}} \cdot \nabla_{\mathbf{p}})\ (\hat{\mathbf{x}} \cdot \mathbf{p}/p_0) = \frac{1-u^2}{p_0}, \tag{62}$$

and if for some k

$$D^k u = \frac{P_k(u)}{p_0^k} \tag{63}$$

where P_k is a constant–coefficient polynomial, then

$$\begin{aligned} D^{k+1} u &= \frac{P_k'(u) Du}{p_0^k} - \frac{k P_k(u)}{p_0^{k+1}} \\ &\equiv \frac{P_{k+1}(u)}{p_0^{k+1}}, \end{aligned} \tag{64}$$

hence eq. (63) holds for $k = 1, 2, \cdots$ by induction. Thus

$$D^k \xi = -\frac{x^0}{R}\, D^k u = -\frac{x^0}{R}\, \frac{P_k(u)}{p_0^k}, \qquad k = 1, 2, \cdots, \tag{65}$$

which implies

$$\big|D^k \xi\big| \leq \frac{1+\epsilon}{m^k} \max_{|u| \leq 1} \big|P_k(u)\big|. \tag{66}$$

But $D^k(\xi^{-1})$ is a polynomial in ξ^{-1} and $D\xi, D^2\xi, \cdots, D^k\xi$; hence eq. (59) follows from eqs. (54) and (66). ∎

The following is an immediate consequence of the above theorem.

Corollary 4.11.

(a) For every $\sigma \in \Sigma$, the form

$$\langle\, f_1 \,|\, f_2 \,\rangle_\sigma = \int_\sigma d\sigma\, \overline{f_1(z)} f_2(z) \tag{67}$$

defines a $\mathcal{P}_0$–invariant inner product on $\mathcal{K}$, under which $\mathcal{K}$ is a Hilbert space.

(b) The transformations $(U_g f)(z) \equiv f(g^{-1}z)$, $g \in \mathcal{P}_0$, form a unitary irreducible representation of $\mathcal{P}_0$ under the above inner product, and the map $\hat{f} \mapsto f$ from $L^2_+(d\tilde{p})$ to $\mathcal{K}$ intertwines this representation with the usual one on $L^2_+(d\tilde{p})$.

(c) For each $\sigma \in \Sigma$, we have the resolution of unity

$$\int_\sigma d\sigma\, |\, e_z \,\rangle\langle\, e_z \,| \,= I \tag{68}$$

on $L^2_+(d\tilde{p})$ (or, equivalently, on $\mathcal{K}$ if e_z is replaced by $\tilde{e}_z$.) ∎

Note: As in section 4.4, all the above results extend by continuity to the case $\lambda = 0$. #

4.6. The Non–Relativistic Limit

We now show that in the non–relativistic limit $c \to \infty$, the foregoing coherent–state representation of $\mathcal{P}_0$ reduces to the representation of $\mathcal{G}_2$ derived in section 4.3, in a certain sense to be made precise. As a by–product, we discover that the Gaussian weight function associated with the latter representation (hence also the closely related weight function associated with the canonical coherent states) has its origin in the geometry of the relativistic (dual) "momentum space" Ω^+_λ. That is, for large $|\mathbf{y}|$ the solutions in $\mathcal{K}$ are dampened by the factor $\exp[-\sqrt{\lambda^2 + \mathbf{y}^2}\omega]$ in momentum space, which in the non–relativistic limit amounts to having a Gaussian weight function in phase space.

In considering the non–relativistic limit, we make all dependence on c explicit but set $\hbar = 1$. Also, it is convenient to choose a coordinate system in which the spacetime metric is $g = \mathrm{diag}(1, -1, \ldots, -1)$,

so that $y^0 = y_0 = \sqrt{\lambda^2 + \mathbf{y}^2}$ and $p^0 = p_0 = \sqrt{m^2c^2 + \mathbf{p}^2}$. Fix $u > 0$ and let $\lambda = uc$. Then

$$\begin{aligned} y_0\omega &= \sqrt{u^2c^2 + \mathbf{y}^2}\sqrt{m^2c^2 + \mathbf{p}^2} \\ &= umc^2 + \frac{u\mathbf{p}^2}{2m} + \frac{m\mathbf{y}^2}{2u} + O(c^{-2}). \end{aligned} \tag{1}$$

Working heuristically at first, we expect that for large c, holomorphic solutions of the Klein–Gordon equation can be approximated by

$$\begin{aligned} f(x - iy) &\equiv \int_{\Omega_m^+} d\tilde{p}\ \exp\left[-it\omega + i\mathbf{x}\cdot\mathbf{p} - y_0\omega + \mathbf{y}\cdot\mathbf{p}\right]\hat{f}(\mathbf{p}) \\ &\sim \int \frac{d^s\mathbf{p}}{(2\pi)^2\cdot 2mc}\ \exp\left[-it\left(mc^2 + \frac{\mathbf{p}^2}{2m}\right) + i\mathbf{x}\cdot\mathbf{p}\right]\times \\ &\qquad \exp\left[-umc^2 - \frac{u\mathbf{p}^2}{2m} - \frac{m\mathbf{y}^2}{2u} + \mathbf{y}\cdot\mathbf{p}\right]\hat{f}(\mathbf{p}) \\ &\sim (2mc)^{-1}\exp\left[-i\tau mc^2 - m\mathbf{y}^2/2u\right]\ f_{\mathrm{NR}}(\mathbf{x} - i\mathbf{y}, \tau), \end{aligned} \tag{2}$$

where $\tau = t - iu$ and f_{NR} is the corresponding holomorphic solution of the Schrödinger equation defined in section 4.3. Note that the Gaussian factor $\exp[-m\mathbf{y}^2/2u]$ is the square root of the weight function for the Galilean coherent states, hence if we choose $\hat{f}(\mathbf{p}) \in L^2(\mathbb{R}^s) \subset L^2_+(d\tilde{p})$, then

$$\|e^{-m\mathbf{y}^2/2u} f_{\mathrm{NR}}\|^2_{L^2(\mathbb{C}^s)} = (\pi u/m)^s\, \|f_{\mathrm{NR}}\|^2_{\mathcal{H}_u} < \infty. \tag{3}$$

We now rigorously justify the above heuristic argument. Let $f(z)$ be the function in $\mathcal{K}$ corresponding to $\hat{f}(\mathbf{p})$ and denote by f_c its restriction to $x^0 = t$ and $y^2 = u^2c^2$, for *fixed* $u > 0$.

Theorem 4.12. *Let $u > 0$ and $\hat{f}(\mathbf{p}) \in L^2(\mathbb{R}^s)$. Then*

$$\begin{aligned} J(c) &\equiv \|2mce^{i\tau mc^2}\, f_c - e^{-m\mathbf{y}^2/2u} f_{\mathrm{NR}}\|^2_{L^2(\mathbb{C}^s)} \\ &\to 0 \text{ as } c \to \infty. \end{aligned} \tag{4}$$

Proof. Without loss of generality, we set $u = m = 1$ and $t = 0$ to simplify the notation. Note first of all that

$$\begin{aligned} \|2ce^{c^2} f_c\|^2_{L^2(\mathbb{C}^s)} &= 4c^2e^{2c^2} A_c \|f\|^2_{\mathcal{K}} \\ &= 4c^2e^{2c^2} A_c \|\hat{f}\|^2_{L^2_+(d\tilde{p})}, \end{aligned} \tag{5}$$

where $A_c \equiv A_\lambda$ ($\lambda \equiv uc = c$). But

$$\begin{aligned} A_c &= \pi^\nu K_{\nu+1}(2c^2) \\ &= (2c)^{-1}\pi^{s/2}e^{-2c^2}\left[1 + O(c^{-2})\right] \end{aligned} \tag{6}$$

and

$$|\hat{f}\|^2_{L^2_+(d\tilde{p})} \leq (2c)^{-1}(2\pi)^{-s}|\hat{f}\|^2_{L^2(\mathbb{R}^s)}. \tag{7}$$

Thus

$$\|2ce^{c^2}f_c\|^2_{L^2(\mathbb{C}^s)} \leq (4\pi)^{-s/2}\|\hat{f}\|^2_{L^2(\mathbb{R}^s)}\left[1 + O(c^{-2})\right], \tag{8}$$

showing that $2ce^{c^2}f_c$ approaches a limit in $L^2(\mathbb{C}^s)$ as $c \to \infty$. Now

$$\begin{aligned} J(c) &= \iint d^s\mathbf{x}\, d^s\mathbf{y} \left| \left[\left(\frac{c}{\omega} e^{c^2 - yp} - e^{-(\mathbf{p}-\mathbf{y})^2/2} \right) \hat{f} \right]^{\vee}(\mathbf{x}) \right|^2 \\ &= \int d^s\mathbf{p} |\hat{f}(\mathbf{p})|^2 \int d^s\mathbf{y} \left(\frac{c}{\omega} e^{c^2 - yp} - e^{-(\mathbf{p}-\mathbf{y})^2/2} \right)^2. \end{aligned} \tag{9}$$

Choose α, γ such that $1/2 < \gamma < \alpha < 1$. Then

$$\begin{aligned} J_1 &\equiv \int_{|\mathbf{p}| > c^{1-\alpha}} d^s\mathbf{p} |\hat{f}(\mathbf{p})|^2 \int_{\mathbb{R}^s} d^s\mathbf{y} \left(\frac{c}{\omega} e^{c^2 - yp} - e^{-(\mathbf{p}-\mathbf{y})^2/2} \right)^2 \\ &\leq 4\pi^{s/2} \|\chi_c \hat{f}\|^2_{L^2(\mathbb{R}^s)} \\ &\to 0 \text{ as } c \to \infty, \end{aligned} \tag{10}$$

where χ_c is the indicator function of the set $\{\mathbf{p} \mid |\mathbf{p}| > c^{1-\alpha}\}$. Define θ and ϕ by $|\mathbf{y}| = c\sinh\theta$ and $|\mathbf{p}| = c\sinh\phi$. Then $y^0 = \cosh\theta$ and $\omega = c^2\cosh\phi$, hence

$$yp \geq c^2\cosh(\theta - \phi) \geq c^2\left[1 + (\theta - \phi)^2/2\right]. \tag{11}$$

Thus for arbitrary $a \geq 0$,

$$\begin{aligned} G_a(\mathbf{p}) &\equiv \int_{|\mathbf{y}|>c\cosh a} d^s\mathbf{y}\, e^{2c^2-2yp} \\ &\le \frac{2c^s\pi^{s/2}}{\Gamma(s/2)} \int_a^\infty d\theta\, \sinh^{s-1}\theta\cosh\theta e^{-c^2(\theta-\phi)^2} \\ &\le \frac{2^{1-s}c^s\pi^{s/.2}}{\Gamma(s/2)} \int_a^\infty d\theta\, e^{(s-1)\theta}\left(e^\theta + e^{-\theta}\right) e^{-c^2(\theta-\phi)^2} \qquad (12) \\ &\le \frac{2^{1-s}c^s\pi^{s/.2}}{\Gamma(s/2)} \int_a^\infty d\theta\, e^{s\theta-c^2(\theta-\phi)^2} \\ &= \frac{2^{1-s}c^s\pi^{s/.2}}{\Gamma(s/2)} e^{s\phi+s^2/4c^2} \int_{c(a-\phi)-s/2c}^\infty du\, e^{-u^2}. \end{aligned}$$

Let $a = \sinh^{-1}(c^{-\gamma})$. Then for $|\mathbf{p}| < c^{1-\alpha}$,

$$c(a-\phi) - s/2c \ge c\left[\sinh^{-1}(c^{-\gamma}) - \sinh^{-1}(c^{-\alpha})\right] - s/2c \equiv g(c). \quad (13)$$

$g(c)$ is independent of $\mathbf{p}$ and $g(c) \sim c^{1-\gamma}$ as $c \to \infty$. Also, $\phi < c^{-\alpha}$ when $|\mathbf{p}| < c^{1-\alpha}$. Hence

$$\begin{aligned} J_2 &\equiv \int_{|\mathbf{p}|<c^{1-\alpha}} d^s\mathbf{p}|\hat{f}(\mathbf{p})|^2 \int_{|\mathbf{y}|>c^{1-\gamma}} d^s\mathbf{y}\, e^{2c^2-2yp} \\ &\le \frac{2^{2-s}c^{s-1}\pi^{s/2}}{\Gamma(s/2)} e^{sc^{-\alpha}+s^2/4c^2} \int_{g(c)}^\infty e^{-u^2} du\, \|\hat{f}\|^2_{L^2(\mathbb{R}^s)} \qquad (14) \\ &\to 0 \text{ as } c \to \infty. \end{aligned}$$

Now

$$\begin{aligned} 2c^2 - 2yp &= y^2 + p^2 - 2yp = (y-p)^2 \\ &= (y_0-\omega)^2/c^2 - (\mathbf{y}-\mathbf{p})^2 \ge -(\mathbf{y}-\mathbf{p})^2. \end{aligned} \qquad (15)$$

Hence

$$\int_{|\mathbf{p}|<c^{1-\alpha}} d^s\mathbf{p}\,|\hat{f}(\mathbf{p})|^2 \int_{|\mathbf{y}|>c^{1-\gamma}} d^s\mathbf{y}\, e^{-(\mathbf{y}-\mathbf{p})^2} \le J_2, \qquad (16)$$

and

$$\int_{|\mathbf{p}|<c^{1-\alpha}} d^s\mathbf{p}\,|\hat{f}(\mathbf{p})|^2 \int_{|\mathbf{y}|>c^{1-\gamma}} d^s\mathbf{y}\,\left(\frac{c}{\omega}e^{c^2-yp}-e^{-(\mathbf{p}-\mathbf{y})^2/2}\right)^2 \leq 4J_2 \to 0 \text{ as } c\to\infty. \tag{17}$$

Finally,

$$\begin{aligned} J_3 &\equiv \int_{|\mathbf{p}|<c^{1-\alpha}} d^s\mathbf{p}\,|\hat{f}(\mathbf{p})|^2 \int_{|\mathbf{y}|<c^{1-\gamma}} d^s\mathbf{y}\,\left(\frac{c}{\omega}e^{c^2-yp}-e^{-(\mathbf{y}-\mathbf{p})^2/2}\right)^2 \\ &= \int_{|\mathbf{p}|<c^{1-\alpha}} d^s\mathbf{p}\,|\hat{f}(\mathbf{p})|^2 \int_{|\mathbf{y}|<c^{1-\gamma}} d^s\mathbf{y}\,e^{-(\mathbf{y}-\mathbf{p})^2}\left(\frac{c}{\omega}e^{c^2\delta^2/2}-1\right)^2, \end{aligned} \tag{18}$$

where

$$\begin{aligned} \delta =& \left|\sqrt{1+\mathbf{y}^2/c^2}-\sqrt{1+\mathbf{p}^2/c^2}\right| \\ &\leq \frac{1}{2c^2}\left|\mathbf{y}^2-\mathbf{p}^2\right| \\ &\leq \frac{1}{2}\left(c^{-2\gamma}+c^{-2\alpha}\right) \\ &\leq c^{-2\gamma}. \end{aligned} \tag{19}$$

We have used the estimate

$$\begin{aligned} \left|\sqrt{1+u^2}-\sqrt{1+v^2}\right| &= \left|\int_u^v \frac{x\,dx}{\sqrt{1+x^2}}\right| \\ &\leq \left|\int_u^v x\,dx\right| = \frac{1}{2}\left|v^2-u^2\right|. \end{aligned} \tag{20}$$

Hence for sufficiently large c and $|\mathbf{p}|<c^{1-\alpha}$,

$$\begin{aligned} \left(\frac{c}{\omega}e^{c^2\delta^2/2}-1\right)^2 &\leq e^{c^2\delta^2}+1-\frac{2c}{\omega}e^{c^2\delta^2/2} \\ &\leq \left(1+2c^2\delta^2\right)+1-2\left(1-\mathbf{p}^2/2c^2\right)e^{c^2\delta^2/2} \\ &\leq 2\left(1-e^{c^2\delta^2/2}\right)+2c^2\delta^2+c^{-2\alpha}e^{c^2\delta^2/2} \\ &\leq 2c^2\delta^2+c^{-2\alpha}\left(1+c^2\delta^2\right) \\ &\equiv h(c)\to 0 \text{ as } c\to\infty. \end{aligned} \tag{21}$$

Thus

$$\begin{aligned} J_3 &\le h(c) \int_{|\mathbf{p}|<c^{1-\alpha}} d^s\mathbf{p}\, |\hat{f}(\mathbf{p})|^2 \int_{|\mathbf{y}|<c^{1-\gamma}} d^s\mathbf{y}\, e^{-(\mathbf{y}-\mathbf{p})^2} \\ &\le h(c)\pi^{s/2}\|\hat{f}\|^2_{L^2(\mathbb{R}^s)} \\ &\to 0 \text{ as } c \to \infty, \end{aligned} \tag{22}$$

which proves that $J(c) \to 0$ as $c \to \infty$. ∎

Notes

This chapter represents the main body of the author's mathematics thesis at the University of Toronto (Kaiser [1977c]). All the theorems, corollaries, lemmas and propositions (labeled 4.1-4.12) have appeared in the literature (Kaiser [1977b, 1978a]). In 1966, when the idea of complex spacetime as a unification of spacetime and phase space first occurred to me, I had found a kind of frame in which *both* the bras and the kets were holomorphic in $\bar{z}$ and the resolution of unity was obtained by a contour integral, using Cauchy's theorem. During a seminar I gave in 1971 at Carleton University in Ottawa (where I was then a post–doctoral fellow in physics), L. Resnick pointed out to me that this "wave–packet representation" appeared to be related to the coherent–state representation, which was at that time unknown to me. The kets were identical to the canonical coherent states, but the bras were not their Riesz duals; in the language of chapter 1, they belonged to a (generalized) frame *reciprocal* to that of the kets, and the resolution of unity was of the type given by eq. (24) in section 1.3, which may be called a continuous version of *biorthogonality*. A version of this result was reported at a conference in Marseille (Kaiser [1974]). I was later informed by J. R. Klauder that a similar representation had been developed by Dirac in connection with quantum electrodynamics (Dirac [1943, 1946]).

The original idea of complex spacetime as phase space was to consider a complex combination of the (symmetric) Lorentzian metric with the (antisymmetric) symplectic structure of phase space, obtaining a hermitian metric on the complex spacetime parametrized by local coordinates of the type $x+ibp$. (I have since learned that this structure, augmented by some technical conditions, is known as a *Kähler*

metric; see Wells [1980].) The above "wave–packet representation" indicated that this combination may in fact be interesting, but so far it was *ad hoc* and lacked a physical basis. Also, the representation was non–relativistic, and it was not at all clear how to extend it to the relativistic domain, as pointed out to me by V. Bargmann in 1975. The standard method of arriving at canonical coherent states is to use an integral transform with a Gaussian kernel in the configuration–space representation, and there is no obvious relativistic candidate for such a kernel. The more general methods described in chapter 3 do not work, since the representations of interest are not square–integrable (section 4.3). An important clue came in 1974 from the study of axiomatic quantum field theory, where I was fascinated by the appearance of tube domains. These domains occur in connection with the analytic continuation of vacuum expectation values of products of fields, and are therefore extensions of such products to complex spcetime. However, the complexified spacetimes themselves are not taken seriously as possible arenas for physics. They are merely used to justify the application of powerful methods from the theory of several complex variables, in order to obtain results concerning the restrictions of vacuum expectaion values to *real* spacetime. (However, the restrictions to *Euclidean* spacetime do have important consequences for statistical mechanics; see Glimm and Jaffe [1981].) I felt that if these tube domains could somehow be given a physical interpretation as extended classical phase spaces, this would give the phase–space formulation of relativistic quantum mechanics a firm physical foundation, since in quantum field theory the extension to complex spacetime is based on solid physical principles such as the spectral condition. This idea was first worked out at the level of non–relativistic quantum mechanics, leading to the representation of the Galilean group given in section 4.3. That amounted to a reformulation of the canonical coherent–state representation in which the Gaussian kernel appears naturally in the *momentum representation,* as a result of the analytic continuation of solutions of the Schrödinger equation. This "explained" the combination $x + ibp$ (section 4.3, eq. (5)) and gave the coherent–state representation a *dynamical* significance. It also cleared the way to the construction of relativistic coherent states, since now the Gaussian kernel merely had to be replaced with the analytic Fourier kernel e^{-izp} on the mass shell. An important tool was the use of groups to compute certain invariant integrals, which I learned from a lecture by E. Stein on Hardy spaces in 1975. The construction of the relativistic coherent states given in sections 4.4 and 4.5 was carried out in 1975–76, culminating in the 1977 thesis. Related results were announced at

a conference in 1976 (Kaiser [1977a]) and at two conferences in 1977 (Kaiser [1977d, 1978b]). To my knowledge, this was the first successful formulation of relativistic coherent states, which have since then gained some popularity (see De Biévre [1989], Ali and Antoine [1989]). An earlier attempt to formulate such states was made by Prugovečki [1976], but this was shown to be inadequate since the proposed states were merely the Gaussian canonical coherent states in disguise, hence not covariant under the Poincaré group (Kaiser [1977c], remark 4 in sec. II.5 and addendum, p. 133.) After the results of the thesis appeared in the literature, Prugovečki [1978; see also 1984] discovered that they can be generalized by replacing the invariant functions e^{-yp} with arbitrary (sufficiently regular) invariant functions. The price of this generalization is that solutions of the Klein–Gordon equation are no longer represented by holomorphic functions and the close connection with quantum field theory (chapter 5) appears to be lost. The relation between the two formalisms and their history was discussed at a conference in Boulder in 1983 (Kaiser [1984b]), where an inconsistency in Prugovečki's formalism was also pointed out.

The classical limit of solutions of the Klein–Gordon equation in the coherent–state representation was studied in Kaiser [1979]. In an effort to understand interactions, the notion of *holomorphic gauge theory* was introduced (Kaiser [1980a, 1981]). This is reviewed in section 6.1. An early attempt was also made to extend the theory to the framework of interacting quantum fields (Kaiser [1980b]), but that was soon abandoned as unsatisfactory. A more promising approach was developed later (Kaiser [1987a]) and is presented in the next chapter.

Note that our phase spaces σ are not unique, since the configuration space S can be chosen arbitrarily as long as it is nowhere timelike and $\lambda > 0$ can be chosen arbitrarily. The freedom in S is, in fact, related to the probability–current conservation, while the freedom in λ, combined with holomorphy, allowed us to express the probability current as a regularization of the usual current by the use of Stokes' theorem (section 4.5, eqs. (37) and (40)). By contrast, the phase spaces obtained by De Biévre [1989] are unique. They are "coadjoint orbits" of the Poincaré group, related to "geometric quantization" theory (Kirillov [1976], Kostant [1970], Souriau [1970]). Although this uniqueness seems attractive, it involves a high cost: the dynamics must be factored out. This means that the ensuing theory is no longer "local in time." Since one of the attractions of coherent–state representations is their "pseudo–locality" in both space and momen-

tum, and since in a relativistic theory time ought to be treated like space, it seems to me an advantage to retain time in the theory. Perhaps a more persuasive argument for this comes from holomorphic gauge theory (section 6.1), where a theory describing a free particle can, in principle, be "perturbed" by introducing a non–trivial fiber metric to obtain a theory describing a particle in an electromagnetic (or Yang–Mills) field. This cannot be done in a natural way once time has been factored out.

Some very interesting work done recently by Unterberger [1988] uses coherent states which are essentially equivalent to ours to develop a pseudodifferential calculus based upon the Poincaré group as an alternative to the usual Weyl calculus, which is based on the Weyl–Heisenberg group. Since the Poincaré group contracts to a group containing the Weyl–Heisenberg group in the non–relativistic limit (section 4.2), Unterberger's "Klein–Gordon calculus" similarly contracts to the Weyl calculus.

Chapter 5
QUANTIZED FIELDS

5.1. Introduction

We have regarded solutions of the Klein–Gordon equation as the quantum states of a relativistic particle. But such solutions also possess another interpretation: they can be viewed as *classical fields,* something like the electromagnetic field (whose components, in fact, satisfy the wave equation, which is the Klein–Gordon equation with zero mass). This interpretation is the basis for quantum field theory. The general idea is that just as the finite number of degrees of freedom of a system of classical particles was quantized to give ordinary ("point") quantum mechanics, a similar prescription can be used to quantize the infinite number of degrees of freedom of a classical field. It turns out that the resulting theory *implies* the existence of particles. In fact, the asymptotic free in– and out–fields are represented by operators which create and destroy particles and antiparticles, in agreement with the fact that such creation and destruction processes occur in nature. These particles and antiparticles are represented by positive–energy solutions of the asymptotic *free* wave equation, e.g. the Klein–Gordon or Dirac equation. Thus the formalism of relativistic quantum mechanics appears to be, at least partially, *absorbed* into quantum field theory.

In regarding solutions of the Klein–Gordon equation as the physical states of a relativistic particle, it was appropriate to restrict our attention to functions having only positive–frequency Fourier components, since the energy of the particle must be positive. Even a small negative energy can be made arbitrarily large and negative by a Lorentz transformation, leading to instability. When the solutions are regarded as classical fields, however, no such restriction on the frequency is necessary or even justifiable. For example, in the case of a *neutral* field (i.e., one not carrying any electric charge), the solutions must be real–valued, hence their Fourier transforms *must* contain negative– as well as positive–frequency components. On the

other hand, the analytic extension of the solutions to complex space-time appeared to depend crucially on the positivity of the energy. We must therefore ask whether an extension is still possible for fields, or if it is even desirable from a physical standpoint, since the connection between solutions and particles is not as immediate as it was earlier. In this chapter we find an affirmative answer to both of these questions. A natural method, which we call the Analytic–Signal transform, will be developed to extend *arbitrary* functions from $\mathbb{R}^{s+1}$ to $\mathbb{C}^{s+1}$, and when the functions represent physical fields, the double tube $\mathcal{T} = \mathcal{T}_+ \cup \mathcal{T}_-$ in $\mathbb{C}^{s+1}$ will be shown to have a direct physical significance as an extended classical phase space, not for the fields themselves but for certain "particle"– and "antiparticle" coherent states $e_z^{\pm}$ associated with them. These states are related directly to the dynamical (interpolating) fields, *not* their asymptotic free in– and out–fields. To be precise, they should be called *charge* coherent states rather than particle coherent states, since they have a well–defined charge whereas, in general, the concept of individual particles does not make sense while interactions are present. If the given fields satisfy some (possibly non–linear) equations, the coherent states satisfy a Klein–Gordon equation with a source term. Hence they represent *dynamical* rather than "bare" particles. For free fields, e_z^+ reduces to the state e_z defined in the last chapter and e_z^- to its complex conjugate, which is a negative–energy solution of the Klein–Gordon equation holomorphic in $\mathcal{T}_-$.

Complex tube domains also appear in the contexts of axiomatic and constructive quantum field theory, and our results suggest that those domains, too, may have interpretations related to classical phase space, a point of view which, to my knowledge, has not been explored heretofore.*

While our extended fields are not analytic in general, they are "analyticity–friendly," i.e. have certain features which yield various analytic objects under different circumstances. For example, their two–point functions are piecewise analytic, and the pieces agree with the analytic Wightman functions. In the special case when the given fields are free, the extended fields themselves are analytic in $\mathcal{T}$. Furthermore, the fields in general possess a *directional* analyticity which

* R. F. Streater has recently told me that G. Källén was informally advocating the interpretation of the holomorphic Wightman two–point function as a correlation function in phase space around 1957. Nothing appears to have been published on this, however.

looks like a covariant version of analyticity in time. Since the latter forms the basis for the continuation of the theory (in the form of vacuum expectaion values) from Lorentzian to Euclidean spacetime (see Nelson [1973a,b] and Glimm and Jaffe [1981]), it may be that our extended fields, when restricted to the Euclidean region, bear some relation to the corresponding Euclidean fields.

The formalism we are about to develop for fields is a natural extension of the one constructed for particles in the last chapter. Like its predecessor, it posesses a degree of regularity not found in the usual spacetime formalism. Some examples of this regularity are:

(a) The extended fields $\phi(z)$ are, under reasonable assumptions, operator–valued *functions* (rather than distributions, as usual) when restricted to $\mathcal{T}$.

(b) The theory contains a natural, covariant ultraviolet damping, which is a permanent feature of the theory. This comes from the possibility of working directly in phase space, away from real spacetime. From the point of view of the usual (*real* spacetime) theory, our formalism looks like a "regularization". From our point of view, however, no regularization is necessary since, it is suggested, *reality takes place in complex spacetime!* In other words, this "regularization" is permanent and is not to be regarded as a kind of trick, used to obtain finite quantities, which must later be removed from the theory.

(c) In the case of free fields, the formalism automatically avoids zero–point energies without normal ordering, due to a *polarization* of the positive– and negative frequency components into the forward and backward tubes, respectively. Observables such as charge, energy–momentum and angular momentum are obtained as conserved integrals of bilinear expressions in the fields over phase spaces $\sigma \subset \mathcal{T}$. These expressions, which are densities for the corresponding observables, look like *regularizations* of the corresponding expressions in the usual spacetime theory. The analytic (Wightman) two–point function acts as a reproducing kernel for the fields, much as it did for the wave functions in chapter 4.

(d) The particles and antiparticles associated with the free Dirac field do not undergo the random motion known as Zitterbewegung (Messiah [1963]), again because of the aforementioned polarization.

5.2. The Multivariate Analytic–Signal Transform

As mentioned above, in dealing with physical *fields* such as the electromagnetic field, rather than quantum *states,* we can no longer justify the restriction that frequencies must be positive. For one thing, as we shall see, in the presence of interactions there is no longer a covariant way to eliminate negative frequencies. Hence the method used in chapter 4 to analytically continue solutions of the Klein–Gordon equation to complex spacetime will no longer work directly. In this section we devise a method for extending *arbitrary* functions from $\mathbb{R}^{s+1}$ to $\mathbb{C}^{s+1}$. When the given functions are positive–energy solutions of the Klein–Gordon or the Wave equation, this method reduces to the analytic extension used in chapter 4. But it is much more general, and will enable us to extend *quantized* fields, whether Bose or Fermi, interacting or free, to complex spacetime. We begin by formulating the method for functions of *one* variable, where it is closely related to the concept of *analytic signals*. For motivational purposes, we think of the variable as time ($s = 0$). In this chapter, Fourier transforms will usually be with respect to *spacetime* ($\mathbb{R}^{s+1}$) rather than just space ($\mathbb{R}^s$). Hence we will denote them by $\tilde{f}$, reserving $\hat{f}$ for the spatial Fourier transform, as done so far.

Suppose we are given a "time–signal," i.e. a real– or complex–valued function of a single real variable x. To begin with, assume that f is a Schwartz test function, although most of our considerations will extend to certain kinds of distributions. Consider the positive– and negative– frequency parts of f, defined by

$$\begin{aligned} f_+(x) &\equiv (2\pi)^{-1} \int_0^\infty dp\, e^{-ixp}\, \tilde{f}(p) \\ f_-(x) &\equiv (2\pi)^{-1} \int_{-\infty}^0 dp\, e^{-ixp}\, \tilde{f}(p). \end{aligned} \tag{1}$$

Then f_+ and f_- extend analytically to the lower–half and upper–half complex planes, respectively, i.e.

$$\begin{aligned} f_+(x - iy) &= (2\pi)^{-1} \int_0^\infty dp\, e^{-i(x-iy)p}\, \tilde{f}(p), \quad y > 0 \\ f_-(x - iy) &= (2\pi)^{-1} \int_{-\infty}^0 dp\, e^{-i(x-iy)p}\, \tilde{f}(p), \quad y < 0. \end{aligned} \tag{2}$$

f_+ and f_- are just the *Fourier–Laplace transforms* of the restrictions of $\tilde{f}$ to the positive and negative frequencies.

If f is complex–valued, then f_+ and f_- are independent and the original signal can be recovered from them as

$$f(x) = \lim_{y \downarrow 0} \left[f_+(x - iy) + f_-(x + iy) \right]. \tag{3}$$

If f is real–valued, then

$$\tilde{f}(p) = \overline{\tilde{f}(-p)}, \tag{4}$$

hence f_+ and f_- are related by reflection,

$$f_+(z) = \overline{f_-(\bar{z})}, \quad z \in \mathbb{C}^-, \tag{5}$$

and

$$f(x) = \lim_{y \downarrow 0} 2\Re f_+(x - iy) = \lim_{y \downarrow 0} 2\Re f_-(x + iy). \tag{6}$$

When f is real, the function $f_+(z)$ is known as the *analytic signal* associated with $f(x)$. A complex–valued signal would have *two* independent associated analytic signals f_+ and f_-. What significance do $f_\pm$ have? For one thing, they are *regularizations* of f. The above equation states that f is jointly a boundary–value of the pair f_+ and f_-. As such, f may actually be quite singular while remaining the boundary–value of analytic functions. Also, $f_\pm$ provide a kind of "envelope" description of f (see Klauder and Sudarshan [1968], section 1.2). For example, if $f(x) = \cos ax$ $(a > 0)$, then $f_\pm(z) = \frac{1}{2} \exp(\mp iaz)$, so the boundary values are $f_\pm(x) = \frac{1}{2} \exp(\mp iax)$.

In order to extend the concept of analytic signals to more than one dimension, let us first of all unify the definitions of f_+ and f_- by defining

$$f(x - iy) \equiv (2\pi)^{-1} \int_{-\infty}^{\infty} dp\, \theta(yp)\, e^{-i(x - iy)p}\, \tilde{f}(p) \tag{7}$$

for *arbitrary* $x - iy \in \mathbb{C}$, where θ is the *unit step function*, defined by

$$\theta(u) = \begin{cases} 0, & u < 0 \\ 1/2, & u = 0 \\ 1, & u > 0. \end{cases} \tag{8}$$

Then we have

$$f(z) = \begin{cases} f_+(z), & y > 0 \\ \frac{1}{2} f(x), & y = 0 \\ f_-(z), & y < 0. \end{cases} \tag{9}$$

[The apparent inconsistency $f(x) = \frac{1}{2} f(x)$ for $y = 0$ is due to a mild abuse of notation. It could be removed by redefining $f(z)$ by a factor of 2 or, more correctly but laboriously, rewriting it as $(Sf)(z)$. We prefer the above notation, since the boundary–values $f(x)$ will not actually be used in the phase–space formalism.]

Let us define the *exponential step function* by

$$\theta^\zeta \equiv \theta(-\Re\zeta)\, e^\zeta, \qquad \zeta \in \mathbb{C}, \tag{10}$$

so that our extension is given by

$$f(z) = (2\pi)^{-1} \int_{-\infty}^{\infty} dp\, \theta^{-izp}\, \tilde{f}(p). \tag{11}$$

The identity

$$\theta(u)\, \theta(u') = \theta(uu')\, \theta(u + u') \tag{12}$$

shows that θ^ζ has the "pseudo–exponential" property

$$\theta^\zeta\, \theta^{\zeta'} = \theta(\Re\zeta\, \Re\zeta')\, \theta^{\zeta+\zeta'}\,, \tag{13}$$

which will be useful later.

Although this unification of f_+ and f_- may at first appear to be somewhat artificial, we shall now see that it is actually very natural. Note first of all that for any real u, we have

$$\theta(u)\, e^{-u} = \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{d\tau}{\tau - i}\, e^{i\tau u}, \tag{14}$$

since the contour on the right–hand side may be closed in the lower half–plane when $u < 0$ and in the upper half–plane when $u > 0$. For $u = 0$, the equation states that

$$\begin{aligned} \theta(0) &= \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{(\tau + i)\, d\tau}{\tau^2 + 1} \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{d\tau}{\tau^2 + 1} = \frac{1}{2}, \end{aligned} \tag{15}$$

in agreement with our definition, if we interpret the integral as the limit as $L \to \infty$ of the integral from $-L$ to L. The exponential step function therefore has the integral representation

$$\theta^{-i(x-iy)p} = \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{d\tau}{\tau - i}\, e^{-i(x-\tau y)p}. \tag{16}$$

If this is substituted into our expression for $f(z)$ and the order of integrations on τ and p is exchanged, we obtain

$$f(x - iy) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{d\tau}{\tau - i}\, f(x - \tau y) \tag{17}$$

for arbitrary $x - iy \in \mathbb{C}$. We shall refer to the right–hand side as the *Analytic–Signal transform* of $f(x)$. It bears a close relation to the *Hilbert transform*, which is defined by

$$(Hf)(x) = \frac{1}{\pi} \mathrm{PV} \int_{-\infty}^{\infty} \frac{du}{u}\, f(x - u), \tag{18}$$

where PV denotes the principal value of the integral. Consider the complex combination

$$\begin{aligned} f(x) - i(Hf)(x) &= \frac{1}{\pi i} \int_{-\infty}^{\infty} du \left[\pi i \delta(u) + \mathrm{PV} \frac{1}{u} \right] f(x - u) \\ &= \frac{1}{\pi i} \lim_{\epsilon \downarrow 0} \int_{-\infty}^{\infty} \frac{du}{u - i\epsilon}\, f(x - u) \\ &= \frac{1}{\pi i} \lim_{\epsilon \downarrow 0} \int_{-\infty}^{\infty} \frac{d\tau}{\tau - i}\, f(x - \tau\epsilon) \\ &= \lim_{\epsilon \downarrow 0} 2 f(x - i\epsilon). \end{aligned} \tag{19}$$

Similarly,

$$f(x) + i(Hf)(x) = \lim_{\epsilon \downarrow 0} 2 f(x + i\epsilon). \tag{20}$$

Hence

$$(Hf)(x) = -i \lim_{\epsilon \downarrow 0} [f(x + i\epsilon) - f(x - i\epsilon)], \tag{21}$$

which, for real–valued f, reduces to

$$(Hf)(x) = \lim_{\epsilon \downarrow 0} 2\Im f(x + i\epsilon) = -\lim_{\epsilon \downarrow 0} 2\Im f(x - i\epsilon). \tag{22}$$

We are now ready to generalize the idea of analytic signals to an arbitrary number of dimensions.

Definition. Let $f \in \mathcal{S}(\mathbb{R}^{s+1})$. The *Analytic–Signal transform* of f is defined by

$$f(x - iy) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{d\tau}{\tau - i} f(x - \tau y). \tag{23}$$

The same argument as above shows that

$$\begin{aligned} f(z) &= (2\pi)^{-s-1} \int_{\mathbb{R}^{s+1}} dp\ \theta^{-izp}\, \tilde{f}(p) \\ &= (2\pi)^{-s-1} \int_{\mathbb{R}^{s+1}} dp\ \theta(yp)\, e^{-ixp-yp}\, \tilde{f}(p). \end{aligned} \tag{24}$$

We shall refer to the right–hand side of this equation as the (inverse) *Fourier–Laplace transform of $\tilde{f}$ in the half–space*

$$M_y \equiv \{p \in \mathbb{R}^{s+1} \,|\, yp \geq 0\}. \tag{25}$$

The integral converges absolutely whenever $\tilde{f} \in L^1(\mathbb{R}^{s+1})$, since $|\theta^{-izp}| \leq 1$. Hence $f(z)$ can actually be defined for some distributions, not only for test functions. The extension of the transform to distributions is complicated by the fact that θ^{-izp} is not a test function in the variable p, hence for a tempered distribution T,

$$T(z) \equiv \tilde{T}(\theta^{-izp}) \tag{26}$$

(defined through the Fourier transform $\tilde{T}$ of T) may not make sense as a function on $\mathbb{C}^{s+1}$. It would be intersting to find a natural class of distributions for which $T(z)$ does make sense as a function. In general, however, it may be necessary to consider distributions T such that $T(z)$ is some kind of *distribution* on $\mathbb{C}^{s+1}$. The solutions to both of these problems are unknown to me, so a certain amount of vagueness will be necessary on this point. In the next section, where T is a quantized field, it will be assumed to satisfy some physically reasonable conditions which imply that $T(z)$ is well–defined in an important *subset* $\mathcal{T}$ of $\mathbb{C}^{s+1}$.

Recall that for $s = 0$, $f(z)$ was analytic in the upper– and lower–half–planes. In more than one dimension, $f(z)$ need not be analytic, even though, for brevity, we still write it as a function of z rather than z and $\bar{z}$. However, $f(z)$ does in general possess a *partial* analyticity which reduces to the above when $s = 0$. Consider the partial derivative of $f(x - iy)$ with respect to $\bar{z}^\mu$, defined by

$$\begin{aligned} 2\bar{\partial}_\mu f(z) &\equiv 2\frac{\partial f}{\partial \bar{z}^\mu} \\ &\equiv \frac{\partial f}{\partial x^\mu} - i\frac{\partial f}{\partial y^\mu}. \end{aligned} \tag{27}$$

Then f is analytic at z if and only if $\bar{\partial}_\mu f = 0$ for all μ. But using our definition of $f(z)$, we find that

$$2\bar{\partial}_\mu f(x - iy) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\tau\, \frac{\partial f}{\partial x^\mu}(x - \tau y). \tag{28}$$

The right–hand side is known as the *X–Ray transform* (Helgason [1984]) of the function $\partial f/\partial x^\mu$, given in terms of the parameters x and y defining the line $x(\tau) = x - \tau y$. It follows that the complex *$\bar{\partial}$–derivative in the direction of y* vanishes, i.e.

$$\begin{aligned} 4\pi y^\mu \bar{\partial}_\mu f(z) &= \int_{-\infty}^{\infty} d\tau\, y^\mu \frac{\partial f}{\partial x^\mu}(x - \tau y) \\ &= -\int_{-\infty}^{\infty} d\tau\, \frac{\partial}{\partial \tau} f(x - \tau y) \\ &= 0, \end{aligned} \tag{29}$$

if f decays for large x (e.g., if f is a test function). Equivalently, using

$$\begin{aligned} 2\bar{\partial}_\mu \left[\theta(yp)\, e^{-izp}\right] &= 2\bar{\partial}_\mu \left[\theta(yp)\right] e^{-izp} \\ &= -i\frac{\partial \theta(yp)}{\partial y^\mu}\, e^{-izp} \\ &= -ip_\mu\, \delta(yp)\, e^{-izp} \\ &= -ip_\mu\, \delta(yp)\, e^{-ixp}, \end{aligned} \tag{30}$$

we have

$$2i\bar{\partial}_\mu f(z) = (2\pi)^{-s-1} \int_{\mathbb{R}^{s+1}} dp\; p_\mu\, \delta(yp)\, e^{-ixp}\, \tilde{f}(p), \tag{31}$$

so $4\pi i\bar{\partial}_\mu f$ is the inverse Fourier transform of $p_\mu \tilde{f}(p)$ in the hyperplane

$$N_y \equiv \{p \in \mathbb{R}^{s+1} \,|\, yp = 0\} = \partial M_y. \tag{32}$$

Hence, if the intersection of the support of $\tilde{f}$ with N_y has positive Lebesgue measure in N_y, then f will not be analytic at $x - iy$ in general. However, in any case,

$$2iy^\mu \bar{\partial}_\mu f(z) = (2\pi)^{-s-1} \int_{\mathbb{R}^{s+1}} dp\ yp\, \delta(yp)\, e^{-ixp}\, \tilde{f}(p) = 0, \tag{33}$$

in agreement with the above conclusion. In the one–dimensional case $s = 0$, this reduces to

$$\frac{\partial f(z)}{\partial \bar{z}} = 0 \quad \forall y \neq 0, \tag{34}$$

which states that $f(z)$ is analytic in the upper– and lower– half–planes. The point is that in one dimension, there are only *two* imaginary directions (up or down), whereas in $s+1$ dimensions, every $y \neq 0$ defines a direction. This motivates the following.

Definition. Let $Y(z)$ be a vector field of type $(0,1)$ on $\mathbb{C}^{s+1}$, i.e. $Y = Y^\mu(z)\,\bar{\partial}_\mu$. Then a function f on $\mathbb{C}^{s+1}$ is *holomorphic along* Y if

$$Yf(z) \equiv Y^\mu(z)\,\bar{\partial}_\mu f(z) \equiv 0. \tag{35}$$

Thus, our Analytic–Signal transform $f(z)$ is holomorphic along $Y(x - iy) = y^\mu\,\bar{\partial}_\mu$.

Note: The "functorial" way to look at $f(z)$ is as an extension of $f(x)$ to the *tangent bundle* $T(\mathbb{R}^{s+1})$. Then the above states that $f(z)$ is holomorphic *along the fibers* of this bundle. It makes sense that $f(z)$ ought to satisfy some constraints, since it is determined by a function $f(x)$ depending on half the number of variables. If $\mathbb{R}^{s+1}$ is replaced by a differentiable manifold, the line $x(\tau) = x - \tau y$ would have to be replaced by a geodesic. This gives a generalized transform which can be used to extend functions on an arbitrary *Riemannian or Pseudo–Riemannian* manifold, such as a curved spacetime, to its tangent bundle. #

The multivariate Analytic–Signal transform is related to the *Hilbert transform in the direction y* (Stein [1970], p. 49), defined as

$$(H_y f)(x) = \frac{1}{\pi}\mathrm{PV}\int_{-\infty}^{\infty}\frac{du}{u}\, f(x-uy), \quad x, y \in \mathbb{R}^{s+1},\ y \neq 0. \tag{36}$$

Namely, an argument similar to the above shows that

$$f(x) + i(H_y f)(x) = \lim_{\epsilon \downarrow 0} 2f(x + i\epsilon y), \tag{37}$$

hence

$$(H_y f)(x) = -i \lim_{\epsilon \downarrow 0} [f(x + i\epsilon y) - f(x - i\epsilon y)], \tag{38}$$

which, for $s = 0$, reduces to the previous relation with the ordinary Hilbert transform. As in the one–dimensional case, $f(x)$ is the boundary–value of $f(z)$ in the sense that

$$f(x) = \lim_{\epsilon \to 0} [f(x + i\epsilon y) + f(x - i\epsilon y)]. \tag{39}$$

For real–valued f, eqs. (38) and (39) reduce to

$$\begin{aligned} f(x) &= \lim_{\epsilon \to 0} 2\Re\, f(x + i\epsilon y) \\ (H_y f)(x) &= \lim_{\epsilon \downarrow 0} 2\Im\, f(x + i\epsilon y). \end{aligned} \tag{40}$$

Note: The Analytic–Signal transform is remarkable in that it combines in a single entity elements of the Hilbert, Fourier–Laplace and X–Ray transforms. In fact, it is an example of a much more general transform which, furthermore, includes the Radon transform and an n–dimensional version of the wavelet transform as special cases! (Kaiser [1990b,c]).

5.3. Axiomatic Field Theory and Particle Phase Spaces

We begin with a study of *general* fields, i.e. fields not necessarily satisfying any differential equation or governed by any particular model of interactions. The Analytic–Signal transform defines (at least *formally*) a canonical extension of such fields to complex spacetime. In this section we show that for general *quantized* fields satisfying the Wightman axioms (Streater and Wightman [1964]), the proposed

extension is natural and interesting. In particular, the *double tube domain* $\mathcal{T} \equiv \mathcal{T}_+ \cup \mathcal{T}_-$ (where $y \in V'_\pm$ in $\mathcal{T}_\pm$) can be interpreted as an extended classical phase space for certain "particle"– and "antiparticle" coherent states naturally associated with the quantized fields. Although the extended fields are in general not analytic (they need not even be *functions*, only ditributions), they are "analyticity–friendly" in the sense that various objects associated with them *are* analytic functions. Consequently, they are more regular than the original fields, which are their boundary values. In particular, we will see that under some reasonable assumtions, the extended fields are operator–valued *functions* (rather than distributions) when restricted to $\mathcal{T}$. The extensions of *free* fields and generalized free fields are, moreover, weakly holomorphic in $\mathcal{T}$.

Recall (chapter 4) that a system with a finite number of degrees of freedom, say a free classical particle, can be quantized in at least two ways: by replacing the position–and momentum variables with operators satisfying the canonical commutation relations (this is called *canonical quantization*), or by considering unitary irreducible representations of the underlying dynamical symmetry group ($\mathcal{G}_2$ for non–relativistic particles and $\mathcal{P}_0$ for relativistic particles). The second scheme seems to be more natural, especially in the relativistic case, where position variables are not a covariant concept. But the first scheme has the advantage that interactions (potentials) can be introduced more easily, since it generates a kind of functional calculus for operators.

Both schemes have counterparts in field quantization. In canonical quantization, the particle position is replaced with the initial configuration of the field ϕ, i.e. with the values of $\phi(x)$ at all space points $\mathbf{x}$ at some initial time $x^0 = t$; its momentum, according to Lagrangian field theory, then corresponds to the time–derivative of the field at time t. Thus the "phase space" of the field is simply the set of all possible initial data on the Cauchy surface $x_0 = t$ in real spacetime. Quantization is implemented by requiring the initial field and its conjugate momentum to satisfy an infinite–dimensional generalization of the canonical commutation relations. Although this is the standard approach to field quantization in the physics literature, its physical and mathematical soundness is open to question. Whereas the two schemes are equivalent when applied to non–relativistic quantum mechanics, it is not clear that they remain equivalent when applied to field quantization, in general. However, they *are* equivalent for the quantization of free fields. We will apply canonical quantization to

the free Klein–Gordon and Dirac fields in sections 5.4 and 5.5.

The second approach to field quantization is subsumed in the so–called axiomatic approach to quantum field theory. The unitary representations of $\mathcal{P}_0$ under which the fields transform are no longer irreducible. This is a consequence of the fact that while $\mathcal{P}_0$ is transitive on the phase space of a classical particle (i.e., any two locations, orientations and states of motion can be transformed into one another), it is no longer transitive on the phase space of a classical field (two sets of initial conditions need not be related by $\mathcal{P}_0$). The reducibility of the representation corresponds to the presence of an infinite set of degrees of freedom or, at the particle level, to an indefinite number of particles. (Even *two* particles would result in reducibility, since the Lorentz norm of their total energy–momentum can be arbitrarily large.) Consequently, the representation is now characterized by an infinite number of parameters and no longer uniquely determines the theory, as it did (for a given mass and spin) when it was irreducible. On the other hand, the label space for the configuration observables, which for N particles in $\mathbb{R}^s$ was $\{1, 2, 3, \cdots, sN\}$, is now $\mathbb{R}^s$, and Lorentz invariance means that the "labels" $\mathbf{x}$ mix with the time variable. This gives an additional structure to quantized fields not shared by ordinary quantum mechanics, and some of this structure is codified in terms of the Wightman axioms, partly making up for the indeterminacy due to the reducibility of the representation.

For simplicity of notation we confine our attention to a single *scalar* field. The results of this section extend to an arbitrary *system* of scalar, spinor or tensor fields. Thus, let $\phi(x)$ be an arbitrary scalar *quantized* field. "Quantized" means that rather than being a real– or complex–valued function on spacetime (like the components of the classical electromagnetic field), $\phi(x)$ is an *operator* on some Hilbert space $\mathcal{H}$. Actually, as noted by Bohr and Rosenfeld [1950], quantized fields are too singular to be measured at a single point in spacetime. This led Wightman to postulate that ϕ is an operator–valued distribution, i.e., when smeared over a test function $f(x)$ it gives an operator $\phi(f)$ (unbounded, in general) on $\mathcal{H}$. The axiomatic approach turns out to provide a surprisingly rich mathematical framework common to all quantum field theories; therefore, it is model–independent. It was followed by *constructive quantum field theory*, in which model field theories are built and shown to satisfy the Wightman axioms. For simplicity, we state the axioms in terms of the formal expression $\phi(x)$ rather than its smeared form $\phi(f)$. Also, we assume, to begin with, that the field ϕ is *neutral,* which means that the operators $\phi(x)$ (or,

rather, their smeared forms $\phi(f)$ with f real–valued) are essentially self–adjoint. Charged fields, which are described by non–Hermitian operators, and spinor (Dirac) fields, will be considered later. The axioms are:

1. *Relativistic Invariance:* Poincaré transformations are implemented by a continuous unitary representation U of $\mathcal{P}_0$ acting on the Hilbert space $\mathcal{H}$ of the theory. (For particles of half–integral spin, such as electrons, $\mathcal{P}_0$ must be replaced by its universal cover; see section 5.5.) Thus,

$$\phi(\Lambda x + a) = U(a, \Lambda)\, \phi(x)\, U(a, \Lambda)^*. \tag{1}$$

   Let P_μ and $M_{\mu\nu}$ be the self–adjoint generators of spacetime translations and Lorentz trnsformations, respectively. They are interpreted physically as the total energy–momentum and angular momentum operators of the field —or, more generally, of the *system* of (possibly coupled) fields. They satisfy the commutation relations of the Lie algebra $\wp$ of $\mathcal{P}_0$. In particular, the P_μ's must commute with one another, thus have a joint spectrum $\Sigma \subset \mathbb{R}^{s+1}$. (Actually, Σ is a subset of the *dual* $(\mathbb{R}^{s+1})^*$ of $\mathbb{R}^{s+1}$, which we may identify with $\mathbb{R}^{s+1}$ using the Minkowski metric; see section 1.1.)

2. *Vacuum:* The Hilbert space contains a unit vector Ψ_0, called the vacuum vector, which is invariant under the representation U, i.e. $U(a, \Lambda)\Psi_0 = \Psi_0$ for all $(a, \Lambda) \in \mathcal{P}_0$. Ψ_0 is unique up to a constant phase factor, and it is *cyclic,* meaning that the set of all vectors of the form

$$\phi(f_1)\, \phi(f_2)\, \cdots \phi(f_n)\Psi_0 \tag{2}$$

   spans $\mathcal{H}$. Invariance under $\mathcal{P}_0$ means that Ψ_0 is a common eigenvector of the generators P_μ and $M_{\mu\nu}$ with eigenvalue zero:

$$\begin{aligned} P_\mu \Psi_0 &= 0 \\ M_{\mu\nu}\Psi_0 &= 0. \end{aligned} \tag{3}$$

3. *Spectral Condition:* The joint spectrum Σ of the energy–momentum operators P_μ is contained in the closed forward light cone:

$$\Sigma \subset \overline{V}_+. \tag{4}$$

   This axiom follows from the physical requirement of *stability,* which merely states that the energy is bounded below. To see

this, note first of all that Σ must be invariant under $\mathcal{L}_0$, by Axiom 1. Hence, if any physical state had a spectral component with $p \notin \overline{V}_+$, that component could be made to have an arbitrarily large negative energy by a Lorentz transformation. Note that the existence and uniqueness of the vacuum means that Σ contains the origin in its point spectrum with multiplicity one.

4. *Locality:* The field operators $\phi(x)$ and $\phi(x')$ at points with spacelike separation commute, i.e.,

$$[\phi(x), \phi(x')] = 0 \qquad \text{if } (x - x')^2 < 0. \tag{5}$$

This corresponds to the physical requirement that measurements of the field at points with spacelike separations must be independent, since no signal can travel faster than light. (For fields of half–integral spin, such as the Dirac field treated in section 5.5, commutators must be replaced with anticommutators.)

5. *Asymptotic Condition:* As the time $x_0 \to \pm\infty$, the field $\phi(x)$ has *weak* asymptotic limits

$$\begin{aligned} \phi(x) &\to \phi_{\text{in}}(x) \quad \text{(weakly) as } x_0 \to -\infty \\ \phi(x) &\to \phi_{\text{out}}(x) \quad \text{(weakly) as } x_0 \to \infty. \end{aligned} \tag{6}$$

ϕ_{in} and ϕ_{out} are *free* fields of mass $m > 0$, i.e. they satisfy the Klein–Gordon equation:

$$(\Box + m^2)\, \phi_{\text{in}} = (\Box + m^2)\, \phi_{\text{out}} = 0. \tag{7}$$

Physically, this means that in the far past and future, all particles are sufficiently far apart to be decoupled, and that ϕ *interpolates* the in– and out– fields. Furthermore, the Hilbert spaces on which $\phi_{\text{in}}, \phi_{\text{out}}$ and ϕ operate all coincide:

$$\mathcal{H}_{\text{in}} = \mathcal{H}_{\text{out}} = \mathcal{H}. \tag{8}$$

In addition to the above, there are axioms concerning the regularity of the distributions $\phi(x)$ (they are assumed to be *tempered*, i.e. the test functions belong to $\mathcal{S}(\mathbb{R}^{s+1})$) and the domains of the smeared operators $\phi(f)$, which we use implicitly, and clustering, which we will not use here.

Let us now draw some conclusions from these axioms. Since the energy–momentum operators generate spacetime translations, it follows from eq. (1) that

$$i\partial_\mu \phi(x) = [\phi(x), P_\mu]. \tag{9}$$

For the Fourier transform $\tilde{\phi}(p)$, this means that

$$[\tilde{\phi}(p), P_\mu] = p_\mu \, \tilde{\phi}(p). \tag{10}$$

Consequently for any $p \in \mathbb{R}^{s+1}$, the "vector"

$$\Phi_p \equiv \tilde{\phi}(-p)\, \Psi_0 \tag{11}$$

(which is in general non–normalizable) satisfies

$$P_\mu \Phi_p = [P_\mu, \tilde{\phi}(-p)]\Psi_0 = p_\mu \Phi_p. \tag{12}$$

That is, Φ_p is a *generalized eigenvector* of energy–momentum with eigenvalue p, unless it vanishes. The spectral condition therefore requires that

$$\tilde{\phi}(-p)\Psi_0 = 0 \qquad \forall p \notin \Sigma_1, \tag{13}$$

where Σ_1 is the intersection of the spectrum Σ with the support of the distribution $\tilde{\phi}$. More generally, if Ψ_p is any generalized eigenvector of energy–momentum with eigenvalue $p \in \Sigma$, then the above commutation relations show that

$$P_\mu \tilde{\phi}(p')\, \Psi_p = (p_\mu - p'_\mu)\, \tilde{\phi}(p')\Psi_p, \tag{14}$$

thus either $\tilde{\phi}(p')\, \Psi_p$ vanishes or it is a generalized eigenvector of energy–momentum with eigenvalue $p - p'$, a necessary condition for which is that $p - p' \in \Sigma$. Thus we conclude that *the operator $\tilde{\phi}(p)$, when it does not vanish, removes an energy–momentum p from the field.* This establishes the physical significance of the field operators, as well as the connection between the (mathematical) Fourier variable p and the (physical) energy–momentum operators P_μ. From the Jacobi identity, it follows that

$$[\,[\tilde{\phi}(p), \tilde{\phi}(p')], P_\mu] = (p_\mu + p'_\mu)\, [\tilde{\phi}(p), \tilde{\phi}(p')], \tag{15}$$

showing that $[\tilde{\phi}(p), \tilde{\phi}(p')]$ (if not zero) removes a total energy–momentum $p + p'$ from the field. The cyclic property of the vacuum, furthermore, implies that the entire spectrum Σ is generated by repeated applications of $\tilde{\phi}$ to the vacuum. Note that in general we

cannot draw any conclusions about the support of $\tilde{\phi}(p)$. For example, $\tilde{\phi}(p)$ need not vanish for spacelike p, since Σ may contain points p' such that $p' - p \in \Sigma$. In fact, very strong conclusions can be drawn from the nature of the support of $\tilde{\phi}$. From Lorentz invariance and $\tilde{\phi}(p)^* = \tilde{\phi}(-p)$ it follows that $\tilde{\phi}(p) = 0$ if and only if $\tilde{\phi}(p') = 0$, where $p' = \pm\Lambda p$ for some $\Lambda \in \mathcal{L}_0$. We conclude that the support of $\tilde{\phi}$ must be a union of sets of the form

$$\begin{aligned} \Omega_m &= \{p \,|\, p^2 = m^2\}, \qquad m > 0 \\ \Omega_0 &= \{p \,|\, p^2 = 0,\ p \neq 0\} \\ \Omega_{00} &= \{0\} \\ \Omega_{im} &= \{p \,|\, p^2 = -m^2\}, \qquad m > 0, \end{aligned} \tag{16}$$

which are, in fact, the various orbits of the *full* Lorentz group $\mathcal{L}$. Greenberg [1962] has shown that ϕ is a *generalized free field* (i.e., a sum or integral of free fields of varying masses $m \geq 0$) if $\tilde{\phi}$ vanishes on any of the following types of sets:

$$\begin{aligned} A &= \Omega_{im}, \qquad m > 0 \\ B &= \Omega_{00} \cup \bigcup_{0 \leq m < M} \Omega_m, \qquad M > 0 \\ C &= \bigcup_{m > M} \Omega_m, \qquad M > 0. \end{aligned} \tag{17}$$

He has also shown, by giving counter–examples, that this conclusion cannot be drawn if $\tilde{\phi}$ vanishes on sets of the type

$$\begin{aligned} D &= \bigcup_{M_1 \leq m < M_2} \Omega_m, \qquad 0 \leq M_1 < M_2 \\ E &= \Omega_{00}. \end{aligned} \tag{18}$$

(See also Dell'Antonio [1961] and Robinson [1962].)

Note: Up to now, *we have not assumed that the field satisfies the canonical commutation relations* (section 5.4), hence our conclusions are quite general and should hold for an arbitrary (system of mutually) interacting field(s). The "Lie algebra" generated by the fields (obtained by including, along with the fields, their commutators

$[\tilde{\phi}(p), \tilde{\phi}(p')]$ as well as higher–order commutators) has a very interesting formal structure, although it is not a Lie algebra in the usual sense. (For one thing, it is uncountably infinite–dimensional rather than finite–dimensional.) Namely, the above relations suggest that the operators P_μ be regarded as belonging to a Cartan subalgebra and that $\tilde{\phi}(p)$ (with p in the support of $\tilde{\phi}$) is a *root vector* with associated root $-p$. The spectrum Σ is therefore reminiscent of a set Δ^+ of positive roots. In general, the Cartan subalgebra consists of a maximal set of commuting observables. When considering charged fields, the charge will also belong to the Cartan subalgebra, with root values $0, \pm\varepsilon, \pm 2\varepsilon, \ldots$, where ε is a fundamental unit of charge. The vacuum is a vector (not in the Lie algebra but in an associated representation space) of "highest" (or lowest) weight which, as in the finite–dimensional case, generates a representation of the algebra because of its cyclic property. To my knowledge, this important analogy between the structures of *general* quantized fields (i.e., apart from the canonical commutation relations or any particular models of interactions) and Lie algebras has not been explored, although the methods of Lie–algebra theory could add a powerful new tool to the study of quantized fields. (In a somewhat different context, the structures of quantized fields and infinite–dimensional Lie algebras are united in string theory; see Green, Schwarz and Witten [1987].) #

Let us now formally extend the quantized field $\phi(x)$ to $\mathbb{C}^{s+1}$, using the Analytic–Signal transform developed in section 5.2. Recall that this transform was originally defined for Schwartz test functions. In principle, we would like to define $\phi(z)$ by using its distributional Fourier transform $\tilde{\phi}(p)$:

$$\begin{aligned} \phi(z) &= (2\pi)^{-s-1} \int_{\mathbb{R}^{s+1}} dp\ \theta^{-izp}\, \tilde{\phi}(p) \\ &\equiv \tilde{\phi}\left(\theta^{-izp}\right). \end{aligned} \tag{19}$$

This presents us with a technical problem, as already noted in the last section, since θ^{-izp} is not a Schwartz test function in p. One way out is to smear $\phi(z)$ with a test function $f(z)$ over $\mathbb{C}^{s+1}$. Although this is the safest solution, it is not very interesting since not much appears to have been gained by extending the field to complex spacetime: the new field is still an operator–valued distribution. However, we shall see that there are reasons to expect $\phi(z)$ to be more regular than a "generic" Analytic–Signal transform, due in part to the fact that $\phi(x)$ satisfies the Wightman axioms. When ϕ is a (generalized) *free* field,

the restriction of $\phi(z)$ to the double tube $\mathcal{T}$ turns out to be a *holomorphic* operator–valued function. We will see that even for general Wightman fields, $\mathcal{T}$ is an important subset of $\mathbb{C}^{s+1}$. In the presence of interactions, holomorphy is lost but some regularity in $\mathcal{T}$ is expected to remain. We now proceed to find conditions which do *not* force ϕ to be a generalized free field but still allow $\phi(z)$ to be an operator–valued *function* on $\mathcal{T}$. The arguments given below have no pretense to rigor; they are only meant to serve as a possible framework for a more precise analysis in the future. All statements and conditions concerning convergence, integrability and decay of operator–valued expressions are meant to hold in the *weak* sense, i.e. for matrix elements between fixed vectors. Since the operators involved are unbounded, we must furthermore assume that the vectors used to form the matrix elements are in their (form) domains.

For a fixed timelike "temper" vector y, θ^{-izp} fails to be a Schwartz test function in two distinct ways: (a) It has a discontinuity on the spacelike hyperplane $N_y = \{p \,|\, yp = 0\}$, and (b) it has a constant modulus on hyperplanes parallel to N_y, hence cannot decay there. On the other hand, by relativistic covariance, the support of $\tilde{\phi}$ must be smeared over the orbits of $\mathcal{L}$, given by eq. (16). This gives a "stratification" of $\tilde{\phi}$ as a sum of tempered distributions

$$\tilde{\phi} = \tilde{\phi}_+ + \tilde{\phi}_0 + \tilde{\phi}_{00} + \tilde{\phi}_- \tag{20}$$

with support properties

$$\begin{aligned} \text{supp } \tilde{\phi}_+ &\subseteq \overline{\bigcup_{m>0} \Omega_m} \equiv \Omega_+ \\ \text{supp } \tilde{\phi}_0 &\subseteq \Omega_0 \\ \text{supp } \tilde{\phi}_{00} &\subseteq \Omega_{00} \\ \text{supp } \tilde{\phi}_- &\subseteq \overline{\bigcup_{m>0} \Omega_{im}} \equiv \Omega_-. \end{aligned} \tag{21}$$

Although Ω_+ and Ω_- contain Ω_0, the distributions $\tilde{\phi}_+$ and $\tilde{\phi}_-$ have no contributions from $p^2 = 0$. (For example, the distribution $1 + \delta(p)$ on $\mathbb{R}$ has a decomposition $T_1 + T_0$ where T_1 has support $\mathbb{R}$ and T_0 has support $\{0\} \subset \mathbb{R}$, but the $p = 0$ contribution to T_1 vanishes.) Similarly, although Ω_0 contains Ω_{00}, $\tilde{\phi}_0$ has no contribution from $p = 0$. Corresponding to the above decomposition of $\tilde{\phi}$, we have, formally,

$$\phi(z) = \phi_+(z) + \phi_0(z) + \phi_{00}(z) + \phi_-(z). \tag{22}$$

We will show that
1. $\phi_+(z)$ and $\phi_0(z)$ are holomorphic operator–valued functions in $\mathcal{T}$;
2. $\phi_{00}(z)$ must be a constant field, to be physically reasonable; and
3. under certain (hopefully not too restrictive) conditions, $\phi_-(z)$, though not holomorphic, is an operator–valued function on $\mathcal{T}$.

First of all, we claim that each of the fields ϕ_α, $(\alpha = \pm, 0, 00)$ is still covariant under $\mathcal{P}_0$.* To see this, take the Fourier transform of eq. (1) and use the invariance of Lebesgue measure dp under $\mathcal{L}_0$. This leads to

$$\tilde{\phi}(p) = e^{iap}\, U(a,\Lambda)\, \tilde{\phi}(\Lambda' p)\, U(a,\Lambda)^*, \tag{23}$$

where Λ' is the transpose of $\Lambda \in \mathcal{L}_0$. Since the different components are "essentially" supported on disjoint subsets and these subsets are invariant under $\mathcal{L}_0$, we conclude that eq. (23) holds for each $\tilde{\phi}_\alpha$. To show that $\phi_+(z)$ and $\phi_0(z)$ are holomorphic in $\mathcal{T}_+$, let $z \in \mathcal{T}_+$. Since θ^{-izp} vanishes for $p \in \overline{V}_- \backslash \{0\}$, and since $\tilde{\phi}_+$ and $\tilde{\phi}_0$ have no contribution from $p = 0$, we have

$$\phi_\alpha(z) = \int_{\overline{V}_+} dp\, e^{-izp}\, \tilde{\phi}_\alpha(p) \quad \alpha = +, 0. \tag{24}$$

Now e^{-izp} may be regarded as the restriction to $\overline{V}_+$ of a Schwartz test function $f_z(p)$ which is of compact support in $\mathbf{p}$ for each fixed p_0 and vanishes when $p_0 < -E$, for some $E > 0$. Thus $\phi_+(z)$ and $\phi_0(z)$ make sense as operator–valued functions on $\mathcal{T}_+$, and they are cleary holomorphic there. (This will be shown explicitly below.) A similar analysis shows that the same can be done for $z \in \mathcal{T}_-$.

Next, we consider $\phi_{00}(z)$. Since $\tilde{\phi}_{00}$ is supported on $\{p = 0\}$, it must have the form $P(\partial)\,\delta(p)$, where $P(\partial)$ is a partial differential operator. In the x–domain (i.e., in real spacetime), this corresponds to a polynomial $P(-ix)$, for which the Analytic–Signal transform is not well–defined, although it is possible that a regularization procedure would cure this. But in any case, non–constant polynomials in x do not appear to be of physical interest since they correspond to

* However, ϕ_α need not satisfy other Wightman axioms such as locality; for example, ϕ_+ need not be a generalized free field.

unbounded fields even in the classical sense (as functions of x). Hence we assume that

(a) $\tilde{\phi}_{00}(p) = 2A\,\delta(p)$, *where* A *is a constant operator.*

This corresponds to a constant field $\phi(x) \equiv 2A$ and, correspondingly, $\phi_{00}(z) \equiv A$ (see eq. (15) in section 5.2). In order that $\phi(z)$ be an operator–valued function on $\mathcal{T}$, it therefore remains only for $\phi_-(z)$ to be one. Note that so far, the only assumption we needed to make, in addition to the Wightman axioms, was *(a)*. To make $\phi_-(z)$ a function, we now make our second assumption:

(b) $\tilde{\phi}_-(p)$ *is integrable on all spacelike hyperplanes. Furthermore, the integral of* $\tilde{\phi}$ *over the hyperplane* $H_{y,\nu} \equiv \{p \,|\, yp = \nu\}$ $(y \in V')$ *grows at most polynomially in* ν.

It is not clear what specific *minimal* conditions on $\tilde{\phi}_-$ produce this property. The integral occurring in *(b)* is known as the *Radon transform* $(R\tilde{\phi})(y,\nu)$ of $\tilde{\phi}$ when y is a (Euclidean) unit vector, and will be further discussed in section 6.2. (See also Helgason [1984].) Unlike the Fourier transform, the Radon transform does not readily generalize to tempered distributions (which were, after all, designed specifically for the Fourier transform). However, it does extend to distributions of compact support and can be further generalized to distributions with only mild decay. Also, the relation of assumptions *(a), (b)* (or their future replacements, if any) to the Wightman axioms needs to be investigated.

In order to compute $\phi_-(z)$ for $z \in \mathcal{T}$ it suffices, by covariance, to do so for $x = 0$ and $y = (u, \mathbf{0})$, for all $u \neq 0$. The analyses for $u > 0$ and $u < 0$ are similar, so we restrict ourselves to $u > 0$. Eq. (19) then gives

$$\phi_-(-iu, \mathbf{0}) = (2\pi)^{-s-1} \int_0^\infty dp_0\, e^{-up_0} \int_{\mathbb{R}^s} d^s\mathbf{p}\, \tilde{\phi}_-(p_0, \mathbf{p}). \qquad (25)$$

For fixed $p_0 \geq 0$, condition *(b)* implies that the integral over $\mathbf{p}$ converges, giving an operator–valued function $F(p_0)$ which is of at most polynomial growth in p_0. $\phi_-(-iu, \mathbf{0})$ is then the Laplace transform of $F(p_0)$, which is indeed well–defined.

Note: The behaviors of $\phi(z)$ and $\tilde{\phi}(p)$ exhibit a certain duality which reflects the dual nature of $y \in \mathbb{R}^{s+1}$ and $p \in (\mathbb{R}^{s+1})^*$ (section 1.1). We have just seen that when $\tilde{\phi}(p)$ behaves reasonably for spacelike p, then $\phi(z)$ behaves reasonably for timelike y. In the trivial case when

$\tilde{\phi}(p) \equiv 0$ for $p^2 < 0$ and *(a)* holds, $\phi(z)$ is holomorphic for $y^2 > 0$. In fact, ϕ is then a generalized free field, hence may be said to be "trivial." This dual behavior also extends to $p^2 \geq 0$ and $y^2 < 0$: For *any* non–constant field, $\tilde{\phi}(p)$ is non–trivial for $p^2 \geq 0$; for spacelike y, the hyperplane $H_{y,\nu}$ (which contains timelike as well as spacelike directions) therefore intersects the support of $\tilde{\phi}$ in a non–compact set and we do not expect $\phi(z)$ to make sense as an operator–valued function outside of $\mathcal{T}$. #

No claims of analyticity can be made for $\phi(z)$ in general. In fact, Greenberg's results show that $\phi(z)$ may not be analytic *anywhere* in $\mathbb{C}^{s+1}$ unless ϕ is a generalized free field. For as in the classical case, formal differentiation with respect to $\bar{z}^\mu$ gives

$$2i\bar{\partial}_\mu \, \phi(z) = (2\pi)^{-s-1} \int_{\mathbb{R}^{s+1}} dp \; p_\mu \, \delta(yp) \, e^{-ixp} \, \tilde{\phi}(p), \qquad (26)$$

hence $4\pi i \bar{\partial}_\mu \, \phi$ is the inverse Fourier transform of $p_\mu \tilde{\phi}$ in the hyperplane N_y. If ϕ is not a generalized free field, then, according to Greenberg, the support of $\tilde{\phi}$ contains sets of timelike as well as sets of spacelike p's with positive Lebesgue measure. Hence, for any nonzero $y \in \mathbb{R}^{s+1}$, the intersection of the support of $\tilde{\phi}$ with N_y has positive measure in N_y, so ϕ will not be holomorphic at $x - iy$ in general. As in the classical case, however, the above equation for $\bar{\partial}_\mu \phi$ implies that $\phi(z)$ is holomorphic along the vector field y, i.e.,

$$y^\mu \bar{\partial}_\mu \phi = 0. \qquad (27)$$

This is a covariant condition which, when specialized to $y = (y_0, \mathbf{0})$, states that $\phi(z)$ is holomorphic in the complex time–direction. As we have seen, this result simply follows from the nature of the Analytic–Signal transform. A similar situation forms the basis of *Euclidean quantum field theory*. However, there one is dealing not directly with the field but with its vacuum expectation values, and the mathematical reason for the analyticity is the spectral condition, which would appear to have little in common with the Analytic–Signal transform.

Incidentally, eq. (26) provides a simple formal proof that $\phi_+(z)$ is holomorphic in $\mathcal{T}$. The support of $\tilde{\phi}_+(p)$ is contained in $\overline{V}$, hence for any y in V', its intersection with N_y is either empty or equal to Ω_{00}. But the contribution from $p = 0$ vanishes, hence eq. (26) shows that $\bar{\partial}_\mu \phi_+(z) = 0$ in $\mathcal{T}$. The same argument also shows that free fields and

generalized free fields are holomorphic in $\mathcal{T}$. In the next two sections we shall study free Klein–Gordon and Dirac fields in more detail.

Although $\phi(z)$ is not holomorphic in general, we will be able to establish for it one essential ingredient of the foregoing phase–space formalism, namely the interpretation of the double tube $\mathcal{T}$ as an extended classical phase space for certain "particles" and "antiparticles" associated with the quantized field ϕ.

First, let us expand the above considerations to include *charged* fields by allowing $\phi(x)$ to be non–Hermitian (i.e., a non–Hermitian operator–valued distribution). Then the extended field $\phi(z)$ need not satisfy the reflection condition $\phi(z)^* = \phi(\bar{z})$. The *charge* Q is defined as a self–adjoint operator which generates overall *phase translations* of the field, i.e.,

$$e^{-i\alpha Q}\,\phi(x)\,e^{i\alpha Q} = e^{i\alpha\varepsilon}\,\phi(x) \tag{28}$$

for real α, where ε is a fundamental unit of charge. This implies

$$\begin{aligned} [\phi(x), Q] &= \varepsilon\,\phi(x) \\ [\tilde{\phi}(p), Q] &= \varepsilon\,\tilde{\phi}(p), \end{aligned} \tag{29}$$

showing that $\phi(x)$ and $\tilde{\phi}(p)$ *remove* a unit ε of charge from the field, while their adjoints *add* a unit of charge. We assume that phase translations commute with Poincaré transformations, and in particular with spacetime translations. Thus

$$[Q, P_\mu] = 0, \tag{30}$$

so charge is *conserved*. Q can be included in the "Cartan subalgebra" containing the P_μ's, and the above commutation relations show that $\tilde{\phi}(p)$ and $\tilde{\phi}(p)^*$ are still "root vectors," with Q–root values $-\varepsilon$ and ε, respectively. We also assume that the vacuum is neutral, i.e. $Q\Psi_0 = 0$. Repeated applications of $\tilde{\phi}$ and $\tilde{\phi}^*$ to Ψ_0 show that the spectrum of Q is $\{0, \pm\varepsilon, \pm 2\varepsilon, \ldots\}$.

Recall that the commutation relation between $\tilde{\phi}(p)$ and P_μ implied that $\tilde{\phi}(p)$ removes an energy–mometum p from the field. Similarly, its adjoint relation

$$[P_\mu, \tilde{\phi}(p)^*] = p_\mu\,\tilde{\phi}(p)^* \tag{31}$$

shows that $\tilde{\phi}(p)^*$ *adds* an energy–momentum p to the field. In place of the generalized eigenvectors Φ_p of energy–momentum which we had for the Hermitian field, we can now define *two* eigenvectors,

$$\begin{aligned} \Phi_p^+ &\equiv \tilde{\phi}(p)^* \, \Psi_0 \\ \Phi_p^- &\equiv \tilde{\phi}(-p) \, \Psi_0, \end{aligned} \tag{32}$$

for each $p \in \Sigma_1$. For a non–Hermitian field, these vectors are independent. They are states of charge ε and $-\varepsilon$, respectively. We may think of them as particles and antiparticles, although they do not have a well–defined mass since p^2 will be variable on Σ_1, unless ϕ is a free field. Each $p \neq 0$ in Σ_1 belongs to the continuous spectrum of the P_μ's, since it can be changed continuously by Lorentz transformations. Hence the "vectors" $\Phi_p^\pm$ are non–normalizable. Since the P_μ's are self–adjoint, and since Φ_p^+ and Φ_p^- belong to different eigenvalues of the charge operator (which is also self–adjoint), we have (with the usual abuse of Dirac notation, where "inner products" of distributions are taken)

$$\begin{aligned} &\langle \, \Phi_p^+ \, | \, \Phi_q^- \, \rangle = 0 \\ &\langle \, \Phi_p^\pm \, | \, \Phi_q^\pm \, \rangle = \sigma(p^2) \, (2\pi)^{s+1} \delta(p-q), \end{aligned} \tag{33}$$

where σ, a distribution with support in Σ_1, depends only on p^2 by Lorentz invariance. (Charge symmetry requires that σ be the same for particles as for antiparticles.) If ϕ is the free field of mass $m > 0$, $\sigma(p^2) = \theta(p_0) \, 2\pi\delta(p^2 - m^2) = 2\pi\delta(p^2 - m^2)$ in $\overline{V}_+ \backslash \{0\}$.

Now define the *particle coherent states* by

$$\begin{aligned} e_z^+ &\equiv \phi(z)^* \Psi_0 \\ &= (2\pi)^{-s-1} \int_{\mathbb{R}^{s+1}} dp \; \theta^{i\bar{z}p} \, \tilde{\phi}(p)^* \, \Psi_0 \\ &= (2\pi)^{-s-1} \int_{\overline{V}_+} dp \; \theta^{i\bar{z}p} \, \Phi_p^+. \end{aligned} \tag{34}$$

Like the Φ_p^+'s, these do not have a well–defined mass; in addition, they are *wave packets,* i.e. have a smeared energy–momentum, but they still have a definite charge ε . Their spectral components are given by

$$\langle \, \Phi_p^+ \, | \, e_z^+ \, \rangle = \sigma(p^2) \, \theta^{i\bar{z}p} = \sigma(p^2) \theta(yp) \, e^{i\bar{z}p}. \tag{35}$$

If z belongs to the backward tube $\mathcal{T}_-$, then $yp < 0$ on $\overline{V}_+\backslash\{0\}$, hence $e_z^+ = 0$. If z belongs to the forward tube $\mathcal{T}_+$, then $yp > 0$ and the vector e_z^+ is weakly holomorphic in $\bar{z}$. For the free field, it reduces to the coherent state e_z defined in chapter 4.

Similarly, define the *coherent antiparticle states* by

$$\begin{aligned} e_z^- &\equiv \phi(z)\Psi_0 \\ &= (2\pi)^{-s-1} \int_{\mathbb{R}^{s+1}} dp\ \theta^{-izp}\tilde{\phi}(p)\,\Psi_0 \\ &= (2\pi)^{-s-1} \int_{\overline{V}_+} dp\ \theta^{izp}\,\Phi_p^-. \end{aligned} \tag{36}$$

These are wave packets of charge $-\varepsilon$ for which

$$\langle\, \Phi_p^- \,|\, e_z^- \,\rangle = \sigma(p^2)\,\theta^{izp} = \sigma(p^2)\,\theta(-yp)\,e^{izp}. \tag{37}$$

Thus e_z^- vanishes in the forward tube and is weakly holomorphic in the backward tube.

In the usual formulation of quantum field theory, particles are associated not directly with the interacting, or interpolating, field ϕ but with its asymptotic fields ϕ_{in} and ϕ_{out}, which are free. (We will construct such free–particle coherent states in the next two sections.) However, the coherent states $e_z^\pm$ *are* directly associated with the interpolating field. We shall refer to them as *interpolating particle coherent states* (section 5.6).

We are now ready to establish the phase–space interpretation of $\mathcal{T}$ in the general case. We will show that $\mathcal{T}_+$ and $\mathcal{T}_-$ are extended phase spaces associated with the particle– and antiparticle coherent states e_z^+ and e_z^-, respectively, in the sense that they parametrize the classical states of these particles.

We first discuss x as a "position" coordinate. In the case of interacting fields there is no hope of finding even a "bad" version of position operators. Recall that position operators were in trouble even in the case of a one–particle theory *without* interactions! In the general case of interacting fields, this problem becomes even more serious, since one is dealing with an indefinite number of particles which may be dynamically created and destroyed. (As argued in section 4.2, the generators M_{0k} of Lorentz boosts qualify as a natural, albeit non–commutative, set of center–of–mass operators; although I believe this idea has merit, it will not be discussed here.) Since no position operators are expected to exist, we must not think of x as eigenvalues or

even expectation values of anything, but rather simply as spacetime parameters or labels.

On the other hand, y will now be shown to be related to the expectations of the energy–momentum operators (which do survive the transition to quantum field theory, as we have seen). For $z, z' \in \mathcal{T}_+$, we have

$$\begin{aligned} \langle\, e^+_{z'} \,|\, e^+_z \,\rangle &= \langle\, \Psi_0 \,|\, \phi(z')\, \phi(z)^* \Psi_0 \,\rangle \\ &= (2\pi)^{-s-1} \int_{\overline{V}_+} dp\; e^{-i(z'-\bar{z})p}\, \sigma(p^2) \\ &= \frac{1}{2\pi} \int_0^\infty dm^2\, \sigma(m^2) \int_{\Omega^+_m} d\tilde{p}\; e^{-i(z'-\bar{z})p} \\ &= \frac{1}{2\pi i} \int_0^\infty dm^2\, \sigma(m^2)\, \Delta^+(z'-\bar{z}; m), \end{aligned} \tag{38}$$

where we have set $m^2 \equiv p^2$ and used

$$(2\pi)^{-s-1}\, dp = (2\pi)^{-s-1}\, dp_0\, d^s\mathbf{p} = (2\pi)^{-1}\, dm^2\, d\tilde{p}, \tag{39}$$

with

$$d\tilde{p} \equiv \left[2(2\pi)^s \sqrt{m^2 + \mathbf{p}^2} \right]^{-1} d^s\mathbf{p} \tag{40}$$

the Lorentz–invariant measure on Ω^+_m. $\Delta^+(w; m)$ is the two–point function for the *free* Klein–Gordon field of mass m, *analytically continued* to $w \equiv z' - \bar{z} \in \mathcal{T}_+$. In the limit $y, y' \to 0$, this gives the *Källén–Lehmann representation* (Itzykson and Zuber [1980]) for the usual two–point function,

$$\langle\, \Psi_0 \,|\, \phi(x')\, \phi(x)^* \Psi_0 \,\rangle = \frac{1}{2\pi i} \int_0^\infty dm^2\, \sigma(m^2)\, \Delta^+(x'-x; m), \tag{41}$$

which is a distribution. In Wightman field theory, such vacuum expectation values are analytically continued using the spectral condition, and conclusions are drawn from these analytic functions about the field in *real* spacetime. In our case, we have *first* extended the field (albeit non–analytically), *then* taken its vacuum expectation values (which, due to the spectral condition, are seen to be analytic functions, not mere distributions). The fact that we arrived at the same

result (i.e., that "the diagram commutes") indicates that our approach is not unrelated to Wightman's. However, there is a fundamental difference: The thesis underlying our work is that the "real" physics actually takes place in *complex* spacetime, and that there is no need to work with the singular limits $y \to 0$.

The norm of e_z^+ is given by

$$\begin{aligned} \|e_z^+\|^2 &= (2\pi)^{-s-1} \int_{\overline{V}_+} dp\ \sigma(p^2)\, e^{-2yp} \\ &= \frac{1}{2\pi} \int_0^\infty dm^2\, \sigma(m^2)\, G(y;m), \end{aligned} \tag{42}$$

where $G(y;m)$, computed in section 4.4, is given by

$$G(y;m) = (2\pi)^{-1} \left(\frac{m}{4\pi\lambda}\right)^\nu K_\nu(2\lambda m). \tag{43}$$

Recall that $\lambda \equiv \sqrt{y^2}$, $\nu \equiv (s-1)/2$ and K_ν is a modified Bessel function. We assume that e_z^+ is normalizable, which means that the spectral density function $\sigma(m^2)$ satisfies the *regularity condition*

$$\begin{aligned} \|e_z^+\|^2 &= (2\pi)^{-2} \int_0^\infty dm^2\, \sigma(m^2)\ \left(\frac{m}{4\pi\lambda}\right)^\nu K_\nu(2\lambda m) \\ &\equiv F(\lambda) < \infty. \end{aligned} \tag{44}$$

(This condition is automatically satisfied for Wightman fields, where it follows from the assumption that ϕ is a tempered distribution; however, it is also satisfied by more singular fields since K_ν decays exponentially.) It follows that

$$\begin{aligned} \langle\, e_z^+ \mid P_\mu\, e_z^+ \,\rangle &= (2\pi)^{-s-1} \int_{\overline{V}_+} dp\ \sigma(p^2)\, p_\mu\, e^{-2yp} \\ &= -\frac{1}{2}\, \frac{\partial F(\lambda)}{\partial y^\mu} \\ &= -y_\mu\, F'(\lambda)/2\lambda. \end{aligned} \tag{45}$$

Using the recursion relation (Abramowitz and Stegun [1964])

$$-\frac{\partial}{\partial\lambda} \left(\lambda^{-\nu}\, K_\nu(2\lambda m)\right) = 2m\lambda^{-\nu} K_{\nu+1}(2\lambda m), \tag{46}$$

we find that the state e_z^+ has an expected energy–momentum

$$\langle\, P_\mu \,\rangle = \left(\frac{m_\lambda}{\lambda}\right) y_\mu, \qquad y \in V_+', \tag{47}$$

where

$$m_\lambda \equiv -\frac{F'(\lambda)}{2F(\lambda)} = \frac{\int_0^\infty dm^2\, \sigma(m^2)\, m^{\nu+1}\, K_{\nu+1}(2\lambda m)}{\int_0^\infty dm^2\, \sigma(m^2)\, m^\nu\, K_\nu(2\lambda m)}. \tag{48}$$

We call m_λ the *effective mass* of the particle coherent states; it generalizes the corresponding quantity for Klein–Gordon particles (section 4.4). The name derives from the relation

$$\langle\, P \,\rangle^2 \equiv \langle\, P_\mu \,\rangle \langle\, P^\mu \,\rangle = m_\lambda^2. \tag{49}$$

It is important to keep in mind that in quantum field theory, the natural picture is the *Heisenberg picture*, where operators evolve in spacetime and states are fixed. Recall that for a free Klein–Gordon particle (chapter 4), we interpreted e_z as a wave packet *focused* about the event $x \equiv \Re z$ and moving with an expected energy–momentum $(m_\lambda/\lambda)\, y$. This suggests that the above states $e_z^\pm$ be given a similar interpretation. Thus z becomes simply a set of *labels* parametrizing the classical states of the particles.

This establishes the interpretation of $\mathcal{T}_+$ as an extended classical phase space associated with the "particle" states e_z^+. A similar computation shows that $\mathcal{T}_-$ acts as an extended classical phase space for the "antiparticle" states e_z^-, whose expected energy–momentum is

$$\langle\, P_\mu \,\rangle = \left(\frac{m_\lambda}{\lambda}\right) (-y_\mu), \qquad y \in V_-'. \tag{50}$$

The expected angular momentum in the states $e_z^\pm$ can be computed similarly. The angular momentum operator $M_{\mu\nu}$ is the generator of rotations in the μ–ν plane, hence

$$i\left(x_\mu \frac{\partial}{\partial x_\nu} - x_\nu \frac{\partial}{\partial x_\mu}\right) \phi(x) = [\phi(x), M_{\mu\nu}]. \tag{51}$$

This implies for the Fourier transform

$$[\tilde{\phi}(p), M_{\mu\nu}] = i\left(p_\mu \frac{\partial}{\partial p^\nu} - p_\nu \frac{\partial}{\partial p^\mu}\right) \tilde{\phi}(p). \tag{52}$$

Since the vacuum is Lorentz–invariant, we have $M_{\mu\nu}\Psi_0 = 0$, hence for $z \in \mathcal{T}_+$,

$$\begin{aligned} M_{\mu\nu}\, e_z^+ &= (2\pi)^{-s-1} \int_{\overline{V}_+} dp\; e^{i\bar{z}p}\, [M_{\mu\nu}, \tilde{\phi}(p)^*]\, \Psi_0 \\ &= i(2\pi)^{-s-1} \int_{\overline{V}_+} dp\; e^{i\bar{z}p} \left(p_\nu \frac{\partial}{\partial p^\mu} - p_\mu \frac{\partial}{\partial p^\nu} \right) \tilde{\phi}(p)^* \Psi_0 \\ &= (2\pi)^{-s-1} \int_{\overline{V}_+} dp\; (\bar{z}_\mu p_\nu - \bar{z}_\nu p_\mu)\; e^{i\bar{z}p}\, \Phi_p^+, \end{aligned} \tag{53}$$

provided that $\tilde{\phi}(p)$ vanishes on the boundary of $\overline{V}_+$ (this excludes massless fields). The expectation of $M_{\mu\nu}$ is therefore related to that of P_μ by

$$\begin{aligned} \langle\, M_{\mu\nu}\,\rangle &= \bar{z}_\mu \langle\, P_\nu\,\rangle - \bar{z}_\nu \langle\, P_\mu\,\rangle \\ &= x_\mu \langle\, P_\nu\,\rangle - x_\nu \langle\, P_\mu\,\rangle \\ &= \left(\frac{m_\lambda}{\lambda}\right)(x_\mu y_\nu - x_\nu y_\mu). \end{aligned} \tag{54}$$

Similarly, in e_z^- with $z \in \mathcal{T}_-$,

$$\langle\, M_{\mu\nu}\,\rangle = -\left(\frac{m_\lambda}{\lambda}\right)(x_\mu y_\nu - x_\nu y_\mu). \tag{55}$$

This section can be summarized by saying that the vector y plays a similar role for general quantized fields as it did for positive–energy solutions of the Klein–Gordon equation, namely it acts as a *control vector* for the energy–momentum. In other words, the function $\theta(yp)\, e^{-yp}$ *acts as a window in momentum space, filtering out from each mass shell* Ω_m *momenta which are not approximately parallel to* y. The step function $\theta(yp)$ makes certain that only parallel components of p pass through this filter by eliminating the antiparallel ones (which would make the integrals diverge). Thus we may think of $\theta(yp)\, e^{-yp}$ as a kind of "ray filter" in $\overline{V}$, when $y \in V'$. We continue to refer to y as a *temper vector* (section 4.4).

Note: The regularity condition given by eq. (44) for $\sigma(m^2)$, i.e. the requirement that e_z^+ be normalizable, shows that λ acts as an effective *ultraviolet cutoff*, since $K_\nu(2\lambda m)$ decays exponentially as $m \to \infty$, giving finite values to m_λ and other quantities associated with the field. #

5.4. Free Klein–Gordon Fields

In the context of general quantum field theory, we were able to show that $\mathcal{T}$ plays the role of an extended phase space for certain "particle" states of the fields. The question arises whether the phase–space formalism of chapter 4 can be generalized to quantized fields. There, we saw that all free–particle states in the Hilbert space could be reconstructed from the values of their wave function on any phase space $\sigma \subset \mathcal{T}_+$. There are two possible ways in which this result might extend to quantized fields: (a) The vectors $e_z^{\pm}$ belong the subspaces $\mathcal{H}_{\pm 1}$ with charge $\pm\varepsilon$, hence we may try to get continuous resolutions, not of the identity on $\mathcal{H}$ but of the orthogonal projection operators $\Pi_{\pm 1}$ to $\mathcal{H}_{\pm 1}$, in terms of these vectors. (This can then be generalized to the resolution of the projection operator Π_n to the subspace $\mathcal{H}_n$ with charge $n\varepsilon$, $n \in \mathbb{Z}$.) (b) The global observables of the theory, such as the energy–momentum, the angular momentum and the charge operators, are usually expressed as conserved integrals of the field operators and their derivatives over an arbitrary configuration space S in spacetime (i.e., an s–dimensional spacelike submanifold of $\mathbb{R}^{s+1}$); our approach would be to express them as integrals of the extended fields over $2s$–dimensional phase spaces σ in $\mathcal{T}$, much as the inner products of positive–energy solutions were expressed as such integrals. In this section we do both of these things for the *free Klein–Gordon field of mass* $m > 0$, which is a quantized solution of

$$(\Box + m^2)\,\phi(x) = 0. \tag{1}$$

We consider classical solutions at first. The Fourier transform $\tilde{\phi}(p)$ has the form

$$\tilde{\phi}(p) = 2\pi\delta(p^2 - m^2)\,a(p) \tag{2}$$

for some complex–valued function $a(p)$ defined on the two–sheeted mass hyperboloid $\Omega_m = \Omega_m^+ \cup \Omega_m^-$. Write

$$b(p) \equiv \overline{a(-p)}, \qquad p \in \Omega_m^+. \tag{3}$$

If the field is neutral, then $\phi(x)$ is real–valued and $b(p) \equiv a(p)$. For charged fields, $a(p)$ and $b(p)$ are independent. At this point, we keep both options open. Then

$$\begin{aligned}\phi(x) &= (2\pi)^{-s} \int_{\mathbb{R}^{s+1}} dp\ \delta(p^2 - m^2)\, e^{-ixp}\, a(p) \\ &= \int_{\Omega_m} d\tilde{p}\ e^{-ixp}\, a(p) \\ &= \int_{\Omega_m^+} d\tilde{p}\ \left[e^{-ixp}\, a(p) + e^{ixp}\, \overline{b(p)}\right].\end{aligned} \tag{4}$$

The extension of $\phi(x)$ to complex spacetime given by the Analytic–Signal transform is

$$\begin{aligned}\phi(x - iy) &= \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{d\tau}{\tau - i}\, \phi(x - \tau y) \\ &= (2\pi)^{-s-1} \int_{\mathbb{R}^{s+1}} dp\ \theta^{-izp}\, \tilde{\phi}(p) \\ &= \int_{\Omega_m} d\tilde{p}\ \theta^{-izp}\, a(p) \\ &= \int_{\Omega_m^+} d\tilde{p}\ \left[\theta(yp)\, e^{-izp}\, a(p) + \theta(-yp)\, e^{izp}\ \overline{b(p)}\right].\end{aligned} \tag{5}$$

If $y \in V_+'$, then $yp > 0$ for all $p \in \Omega_m^+$, hence

$$\phi(z) = \int_{\Omega_m^+} d\tilde{p}\ e^{-izp}\, a(p) \tag{6}$$

is analytic in $\mathcal{T}_+$, containing only the positive–frequency part of the field. Similarly, when $y \in V_-'$, then $yp < 0$ for all $p \in \Omega_m^+$ and

$$\phi(z) = \int_{\Omega_m^+} d\tilde{p}\ e^{izp}\, \overline{b(p)}, \quad z \in \mathcal{T}_-. \tag{7}$$

Thus $\phi(z)$ is also analytic in $\mathcal{T}_-$, where it contains only the negative–frequency part of the field. However, note that the two domains of analyticity $\mathcal{T}_+$ and $\mathcal{T}_-$ do not intersect, hence the corresponding restrictions of $\phi(z)$ need not be analytic continuations of one another.

We are now ready to quantize $\phi(z)$. This will be done by first quantizing $\phi(x)$ and then using the Analytic–Signal transform to extend it to complex spacetime. We assume, to begin with, that ϕ is a

neutral field, so $b(p) \equiv a(p)$. According to the standard rules (Itzykson and Zuber [1980]) of field quantization, $\phi(x)$ becomes an operator on a Hilbert space $\mathcal{H}$ such that at any fixed time x_0, the field "configuration" operators $\phi(x_0, \mathbf{x})$ and their conjugate "momenta" $\partial_0\phi(x_0, \mathbf{x})$ obey the *equal–time commutation relations*

$$\begin{aligned} [\phi(x), \phi(x')]_{x'_0=x_0} &= 0 \\ [\phi(x), \partial_0\phi(x')]_{x'_0=x_0} &= i\delta(\mathbf{x} - \mathbf{x}'). \end{aligned} \tag{8}$$

This is an extension to infinite degrees of freedom of the canonical commutation relations obeyed by the quantum–mechanical position– and momentum operators. Note that since time evolution is to be implemented by a unitary operator, the same commutation relations will then hold at any other time as well. For the non–Hermitian operators $a(p)$, the corresponding commutation relations are

$$[a(p), a(p')] = 2p_0\,\theta(-p_0p'_0)\,(2\pi)^s\,\delta(\mathbf{p} + \mathbf{p}') \tag{9}$$

for $p, p' \in \Omega_m$. Using the neutrality condition $a(-p) = a(p)^*$, these can be rewritten in their conventional form

$$\begin{aligned} [a(p), a(p')] &= 0 \\ [a(p), a(p')^*] &= 2\omega\,(2\pi)^s\,\delta(\mathbf{p} - \mathbf{p}') \end{aligned} \tag{10}$$

where now $p, p' \in \Omega_m^+$.

A charged field can be built up from a pair of neutral fields as

$$\phi(x) = \frac{\phi_1(x) + i\phi_2(x)}{\sqrt{2}}, \tag{11}$$

where the two fields ϕ_1, ϕ_2 each obey the equal–time commutation relations and commute with one another. Equivalently, the operators $a(p)$ and $b(p)$ become independent and satisfy

$$\begin{aligned} [a(p), a(p')] = [b(p), b(p')] = [a(p), b(p')] = 0, \\ [a(p), a^*(p')] = [b(p), b^*(p')] = 2\omega\,(2\pi)^s\,\delta(\mathbf{p} - \mathbf{p}') \end{aligned} \tag{12}$$

for $p, p' \in \Omega_m^+$. The canonical commutation relations for both neutral and charged fields can be put in the manifestly covariant form

$$\begin{aligned}[\tilde{\phi}(p), \tilde{\phi}(p')^*] &= 4\pi^2\, \delta(p^2 - m^2)\, \delta(p'^2 - m^2)\, [a(p), a(p')^*] \\ &= (2\pi)^{s+2} 2p_0\, \theta(p_0 p_0')\, \delta(p^2 - m^2)\, \delta(p'^2 - p^2)\, \delta(\mathbf{p} - \mathbf{p}') \\ &= (2\pi)^{s+2} 2p_0\, \theta(p_0 p_0')\, \delta(p^2 - m^2)\, \delta(p_0'^2 - p_0^2)\, \delta(\mathbf{p} - \mathbf{p}') \\ &= \mathrm{sign}(p_0)\, (2\pi)^{s+2}\, \delta(p^2 - m^2)\, \delta(p - p') \end{aligned} \tag{13}$$

for arbitrary $p, p' \in \mathbb{R}^{s+1}$. For charged fields, this must be supplemented by

$$[\tilde{\phi}(p), \tilde{\phi}(p')] = 0, \qquad p, p' \in \mathbb{R}^{s+1}. \tag{14}$$

Recall that in the general case we had

$$\langle\, \Phi_p^+ \,|\, \Phi_{p'}^+ \,\rangle = \sigma(p^2)\, (2\pi)^{s+1}\, \delta(p - p') \tag{15}$$

for $p, p' \in \overline{V}_+$, where $\sigma(p^2)$ is the spectral density for the two–point function (sec. 5.3, eq. (33)). For the free field now under consideration we have

$$\begin{aligned}\langle\, \Phi_p^+ \,|\, \Phi_{p'}^+ \,\rangle &= \langle\, \Psi_0 \,|\, \tilde{\phi}(p)\, \tilde{\phi}(p')^*\, \Psi_0 \,\rangle \\ &= \langle\, \Psi_0 \,|\, [\tilde{\phi}(p)\,, \tilde{\phi}(p')^*]\, \Psi_0 \,\rangle \\ &= (2\pi)^{s+2}\, \delta(p^2 - m^2)\, \delta(p - p'), \end{aligned} \tag{16}$$

which shows that the spectral density for the free field is

$$\sigma(p^2) = 2\pi\, \delta(p^2 - m^2). \tag{17}$$

The spectral condition implies that

$$a(p)\Psi_0 = 0, \qquad b(p)\Psi_0 = 0 \qquad \forall p \in \Omega_m^+, \tag{18}$$

since otherwise these would be states of energy–momentum $-p$. Hence the particle coherent states defined in the last section are now given by

$$\begin{aligned} e_z^+ \equiv \phi(z)^*\, \Psi_0 &= \int_{\Omega_m^+} d\tilde{p}\; e^{i\bar{z}p}\, a(p)^*\, \Psi_0 \\ &\equiv \int_{\Omega_m^+} d\tilde{p}\; e^{i\bar{z}p}\, \tilde{\Phi}_p^+, \qquad z \in \mathcal{T}_+, \end{aligned} \tag{19}$$

where the vectors $\tilde{\Phi}_p^+$ are generalized eigenvectors of energy–momentum $p \in \Omega_m^+$ with the normalization

$$\begin{aligned} \langle\, \tilde{\Phi}_p^+ \,|\, \tilde{\Phi}_{p'}^+ \,\rangle &= \langle\, \Psi_0 \,|\, a(p)\, a(p')^* \Psi_0 \,\rangle \\ &= \langle\, \Psi_0 \,|\, [a(p),\, a(p')^*] \Psi_0 \,\rangle \\ &= 2\omega (2\pi)^s\, \delta(\mathbf{p} - \mathbf{p}'). \end{aligned} \tag{20}$$

The wave packets e_z^+ span the one–particle subspace $\mathcal{H}_1$ of $\mathcal{H}$ and have the momentum representation

$$\langle\, \tilde{\Phi}_p^+ \,|\, e_z^+ \,\rangle = e^{i\bar{z}p}. \tag{21}$$

A dense subspace of $\mathcal{H}_1$ is obtained by applying the smeared operators

$$\phi^*(f) \equiv \phi(\bar{f})^* \equiv \int dx\, \phi(x)^*\, f(x) = (2\pi)^{-s-1} \int_{\mathbb{R}^{s+1}} dp\;\; \tilde{\phi}(p)^*\, \tilde{f}(p) \tag{22}$$

to the vacuum, where f is a test function. This gives

$$\begin{aligned} \Psi_f^+ \equiv \phi^*(f)\, \Psi_0 &= (2\pi)^{-s-1} \int_{\mathbb{R}^{s+1}} dp\;\, \tilde{f}(p)\, \tilde{\phi}^*(p)\, \Psi_0 \\ &= (2\pi)^{-s-1} \int_{\mathbb{R}^{s+1}} dp\;\, \tilde{f}(p)\, \Phi_p^+ = \int_{\Omega_m^+} d\tilde{p}\;\, \hat{f}(p)\, \tilde{\Phi}_p^+, \end{aligned} \tag{23}$$

where $\hat{f}$ is the restriction of $\tilde{f}$ to Ω_m^+. Hence the functions

$$\langle\, e_z^+ \,|\, \Psi_f^+ \,\rangle = \int_{\Omega_m^+} d\tilde{p}\; e^{-izp}\, \hat{f}(p) \tag{24}$$

are exactly the holomorphic positive–energy solutions of the Klein–Gordon equation discussed in section 4.4, with e_z^+ corresponding to the evaluation maps e_z. The space $\mathcal{K}$ of these solutions can thus be identified with $\mathcal{H}_1$, and the orthogonal projection from $\mathcal{H}$ to $\mathcal{H}_1$ is given by

$$(\Pi_1 \Psi)(z) = \langle\, e_z^+ \,|\, \Psi \,\rangle. \tag{25}$$

Consequently, the resolution of unity developed in chapter 4 can now be restated as a resolution of Π_1:

$$\Pi_1 = \int_{\sigma_+} d\sigma \; | \, e_z^+ \, \rangle \langle \, e_z^+ \, | \, , \tag{26}$$

where σ_+, earlier denoted by σ, is a *particle phase space*, i.e. has the form

$$\sigma_+ = \{x - iy \, | \, x \in S, \; y \in \Omega_\lambda^+\} \tag{27}$$

for some $\lambda > 0$ and some spacelike or, more generally, nowhere timelike (see section 4.5) submanifold S of real spacetime. As in section 4.5, the measure $d\sigma$ is given in terms of the Poincaré–invariant symplectic form $\alpha = dy_\mu \wedge dx^\mu$ by restricting $\alpha^s \equiv \alpha \wedge \cdots \wedge \alpha$ to σ_+ and choosing an orientation:

$$d\sigma = (s! A_\lambda)^{-1} \, \alpha^s = A_\lambda^{-1} \, \widehat{dy}^{\,\mu} \wedge \widehat{dx}_\mu . \tag{28}$$

Similarly, the antiparticle coherent states for the free field are given by

$$\begin{aligned} e_z^- \equiv \phi(z) \, \Psi_0 &= \int_{\Omega_m^+} d\tilde{p} \; e^{izp} \, b(p)^* \, \Psi_0 \\ &\equiv \int_{\Omega_m^+} d\tilde{p} \; e^{izp} \, \tilde{\Phi}_p^- , \qquad z \in \mathcal{T}_- . \end{aligned} \tag{29}$$

Since for $p \in \Omega_m^+$ and $z \in \mathcal{T}_-$ we have

$$\langle \, \tilde{\Phi}_p^- \, | \, e_z^- \, \rangle = e^{izp} = \langle \, \tilde{\Phi}_p^+ \, | \, e_{\bar{z}}^+ \, \rangle, \tag{30}$$

it follows that e_z^- has exactly the same spacetime behavior as $e_{\bar{z}}^+$, confirming the interpretation of an antiparticle as a particle moving backward in time. An *antiparticle phase space* is defined as a submanifold of $\mathcal{T}_-$ given by

$$\sigma_- = \{x - iy \, | \, x \in S, \; y \in \Omega_\lambda^-\}, \tag{31}$$

where S is as above. The resolution of Π_{-1} is then given by

$$\Pi_{-1} = \int_{\sigma_-} d\sigma \; | \, e_z^- \, \rangle \langle \, e_z^- \, | \, . \tag{32}$$

Many–particle or –antiparticle coherent states and their corresponding phase spaces can be defined similarly, and the commutation

relations imply that such states are *symmetric* with respect to permutations of the particles' complex coordinates. For example,

$$e^+_{z_1 z_2} \equiv \phi(z_1)^* \phi(z_2)^* \Psi_0 = e^+_{z_2 z_1},$$

since $\phi(z_1)$ and $\phi(z_2)$ commute. In this way, a phase–space formalism can be buit for an indefinite number of particles (or charges), analogous to the grand–canonical ensemble in classical statistical mechanics. This idea will not be further pursued here. Instead, we now embark on option (b) above, i.e. the construction of global, conserved field observables as integrals over particle and antiparticle phase spaces.

The *particle number* and *antiparticle number* operators are given by

$$\begin{aligned} N_+ &= \int_{\Omega^+_m} d\tilde{p}\ a(p)^*\, a(p) \\ N_- &= \int_{\Omega^+_m} d\tilde{p}\ b(p)^*\, b(p), \end{aligned} \tag{33}$$

N_+ and N_- generalize the harmonic–oscillator Hamiltonian A^*A to the infinite number of degrees of freedom possessed by the field, where normal modes of vibration are labeled by $p \in \Omega^+_m$ for particles and $p \in \Omega^-_m$ for antiparticles. The total charge operator is $Q = \varepsilon\,(N_+ - N_-)$, as can be seen from its commutation relations with $a(p)$ and $b(p)$. But the resolution of unity derived in chapter 4 can now be restated as

$$\begin{aligned} \int_{\sigma_+} d\sigma\ \exp(i\bar{z}p - izp') &= (2\pi)^s\, 2\omega(\mathbf{p})\, \delta(\mathbf{p}-\mathbf{p}') = \langle\, \tilde{\Phi}^+_p \mid \tilde{\Phi}^+_{p'} \,\rangle \\ \int_{\sigma_-} d\sigma\ \exp(izp - i\bar{z}p') &= (2\pi)^s\, 2\omega(\mathbf{p})\, \delta(\mathbf{p}-\mathbf{p}') = \langle\, \tilde{\Phi}^-_p \mid \tilde{\Phi}^-_{p'} \,\rangle \end{aligned} \tag{34}$$

for $p, p' \in \Omega^+_m$, where the second identity follows from the first by replacing z with $\bar{z}$ and σ_+ with σ_-. It follows that $N_\pm$ can be expressed as phase–space integrals of the extended field $\phi(z)$:

$$\begin{aligned} N_+ &= \int_{\sigma_+} d\sigma\, \phi(z)^*\, \phi(z) \\ N_- &= \int_{\sigma_-} d\sigma\, \phi(z)\, \phi(z)^*. \end{aligned} \tag{35}$$

Hence the charge is given by

$$Q = \varepsilon \int_{\sigma_+} d\sigma\, \phi(z)^*\, \phi(z) - \varepsilon \int_{\sigma_-} d\sigma\, \phi(z)\, \phi(z)^*. \tag{36}$$

The two integrals can be combined into one as follows: Define the *total* phase space as $\sigma = \sigma_+ - \sigma_-$, where the minus sign means that σ_- enters with the *opposite* ("negative") orientation to that of σ_+, in the sense of *chains* (Warner [1971]). The reason for this choice of orientation is that B_λ^+ and B_λ^- are both open sets of $\mathbb{R}^{s+1}$, hence must have the same orientation, and we orient Ω_λ^+ and Ω_λ^- so that

$$\Omega_\lambda^\pm = -\partial\, B_\lambda^\pm. \tag{37}$$

Since the outward normal of B_λ^+ points "down" and that of B_λ^- points "up," this means that Ω_λ^- must have the opposite orientation to that of Ω_λ^+. Thus, setting $B_\lambda \equiv B_\lambda^+ + B_\lambda^-$ and $\Omega_\lambda \equiv \Omega_\lambda^+ - \Omega_\lambda^-$, we have

$$\Omega_\lambda = -\partial B_\lambda. \tag{38}$$

This gives σ_- the orientation opposite to that of σ_+, and we have

$$\sigma \equiv S \times \Omega_\lambda = \sigma_+ - \sigma_-. \tag{39}$$

Next, define the *Wick–ordered product* (or normal product) by

$$:\phi(z)^*\, \phi(z):\equiv \begin{cases} \phi(z)^*\, \phi(z), & z \in \mathcal{T}_+ \\ \phi(z)\, \phi(z)^*, & z \in \mathcal{T}_-. \end{cases} \tag{40}$$

This coincides with the usual definition, since in $\mathcal{T}_+$, ϕ^* is a creation operator and ϕ is an annihilation operator, while in $\mathcal{T}_-$ these roles are reversed. The charge can now be written in the compact form

$$Q = \varepsilon \int_\sigma d\sigma\ :\phi(z)^*\, \phi(z): \tag{41}$$

We may therefore interpret the operator

$$\rho(z) \equiv \varepsilon :\phi(z)^*\, \phi(z): \tag{42}$$

as a *scalar phase–space charge density* with respect to the measure $d\sigma$.

The Wick ordering can be viewed as a special case of *imaginary–time ordering*, if we define $\phi^*(z) \equiv \phi(\bar{z})^*$:

$$: \phi(z')^* \, \phi(z) :\equiv: \phi^*(\bar{z}') \, \phi(z) := T_i \left[\phi^*(\bar{z}') \, \phi(z) \right], \tag{43}$$

where

$$\begin{aligned} & T_i \left[\phi^*(z') \, \phi(z) \right] \\ & \quad \equiv \theta \left(\Im(z_0' - z_0) \right) \, \phi^*(z') \, \phi(z) + \theta \left(\Im(z_0 - z_0') \right) \, \phi(z) \, \phi^*(z') \end{aligned} \tag{44}$$

for $z, z' \in \mathcal{T}$. This definition is Lorentz–invariant, since

$$T_i[\phi(z') \, \phi(z)] = \phi(z') \, \phi(z) = \phi(z) \, \phi(z') \quad \forall z, z' \in \mathcal{T} \tag{45}$$

and

$$T_i[\phi^*(z') \, \phi(z)] = \phi^*(z') \, \phi(z) = \phi(z) \, \phi^*(z') \tag{46}$$

when z and z' are in the same half of $\mathcal{T}$, whereas if they are in opposite halves of $\mathcal{T}$, the sign of $\Im(z_0' - z_0)$ is invariant.

Note: For the extended fields, the Wick ordering is not a necessity but a mere convenience, allowing us to combine the integrals over σ_+ and σ_- into a single integral. Each of these integrals is *already* in normal order, since the extension to complex spacetime *polarizes* the free field into its positive–and negative–frequency parts. Also, the extended fields are operator–valued *functions* rather than distributions, hence products such as $\phi(z)^* \phi(z)$ are well–defined, which is *not* the case in the usual formalism. A similar situation will occur in the expressions for the other observables (energy–momentum, angular momentum, etc.) as phase–space integrals. Hence *the phase–space formalism resolves the problem of zero–point energies* without the need to subtract infinite terms "by hand"! In this connection, see the remarks on p. 21 of Henley and Thirring [1962]. #

The above expression for the charge can be related to the usual one in the spacetime formalism, which is

$$\begin{aligned} Q_{\text{usual}} &= i\varepsilon \int_S \widehat{dx}_\mu \; : \phi^* \frac{\partial \phi}{\partial x_\mu} - \frac{\partial \phi^*}{\partial x_\mu} \cdot \phi(x) : \\ &\equiv \int_S \widehat{dx}_\mu \, J^\mu(x), \end{aligned} \tag{47}$$

by using $\Omega_\lambda = -\partial B_\lambda$ and applying Stokes' theorem:

$$\begin{aligned} Q &= \varepsilon\, A_\lambda^{-1} \int_S \widehat{dx}_\mu \int_{\Omega_\lambda} \widehat{dy}^{\,\mu} : \phi^*\, \phi : \\ &= -\varepsilon\, A_\lambda^{-1} \int_S \widehat{dx}_\mu \int_{B_\lambda} dy\, \frac{\partial}{\partial y_\mu} : \phi^*\, \phi : \\ &\equiv A_\lambda^{-1} \int_S \widehat{dx}_\mu \int_{B_\lambda} dy\, j^\mu(x - iy), \end{aligned} \tag{48}$$

where

$$j^\mu(z) \equiv -\frac{\partial}{\partial y_\mu}\, \rho(z) \tag{49}$$

is the *phase–space current density*. Using the notation

$$\partial^\mu \equiv \frac{\partial}{\partial z_\mu} \equiv \frac{1}{2}\left(\frac{\partial}{\partial x_\mu} + i\frac{\partial}{\partial y_\mu}\right), \tag{50}$$

we have

$$-\frac{\partial}{\partial y_\mu} = i(\partial^\mu - \bar\partial^\mu). \tag{51}$$

Hence, by the holomorphy of ϕ,

$$\begin{aligned} j^\mu(z) &\equiv -\varepsilon\, \frac{\partial}{\partial y_\mu} : \phi^*\, \phi : \\ &= i\varepsilon\, \left(\partial^\mu - \bar\partial^\mu\right) : \phi^*\, \phi : \\ &= i\varepsilon : \phi^*\, \partial^\mu \phi - \bar\partial^\mu \phi^* \cdot \phi : \\ &= i\varepsilon : \phi^* \frac{\partial \phi}{\partial x^\mu} - \frac{\partial \phi^*}{\partial x^\mu}\, \phi : . \end{aligned} \tag{52}$$

Our expression for the charge is therefore

$$Q = \int_S \widehat{dx}_\mu\, J^\mu_{(\lambda)}(x), \tag{53}$$

where

$$\begin{aligned} J^\mu_{(\lambda)}(x) &\equiv A_\lambda^{-1} \int_{B_\lambda} dy\, j^\mu(x - iy) \\ &= i\varepsilon\, A_\lambda^{-1} \int_{B_\lambda} dy : \phi^* \frac{\partial \phi}{\partial x^\mu} - \frac{\partial \phi^*}{\partial x^\mu}\, \phi : \end{aligned} \tag{54}$$

is seen to be a *regularized version* of the usual current density $J^\mu(x)$ obtained by first extending it to $\mathcal{T}$ and then integrating it over B_λ.

The vector field $j^\mu(z)$ is conserved in *real* spacetime for each fixed $y \in V'$, since

$$\begin{aligned}\frac{\partial j^\mu}{\partial x^\mu} &= -\varepsilon \, \frac{\partial^2}{\partial x^\mu \partial y_\mu} : \phi^* \, \phi : \\ &= i\varepsilon \, (\partial_\mu + \bar{\partial}_\mu)\,(\partial^\mu - \bar{\partial}^\mu) : \phi^* \, \phi : \\ &= i\varepsilon \, (\Box_z - \Box_{\bar{z}}) : \phi^* \, \phi : \\ &= i\varepsilon : (\phi^* \, \Box_z \phi - \Box_{\bar{z}} \phi^* \cdot \phi) : \\ &= 0,\end{aligned} \tag{55}$$

by virtue of the Klein–Gordon equation combined with the holomorphy of ϕ in $\mathcal{T}$. This implies that $J^\mu_{(\lambda)}(x)$ is also conserved, hence the charge does not depend on the choice of S or σ.

Note: In using Stokes' theorem above, we have assumed that the contribution from $|y_0| \to \infty$ vanishes. (This was implicit in writing the non–compact manifold Ω_λ as $-\partial B_\lambda$.) This is indeed the case, as has been shown rigorously in the context of the one–particle theory in chapter 4 (theorem 4.10). Also, we see another example of the pattern, mentioned before, that in the phase–space formalism vector– and tensor fields can often be derived from *scalar potentials.* Here, $\rho(z)$ acts as a potential for $j^\mu(z)$. Note also that the Klein–Gordon equation can be written in the form

$$(\Box_z + m^2)\rho(z) = \varepsilon : \phi^*(\Box_z + m^2)\phi := 0, \tag{56}$$

which is manifestly gauge–invariant. #

Recall now that $\phi(z)$ is a "root vector" of the charge with root value $-\varepsilon$:

$$[\phi(z'), Q] = \varepsilon \, \phi(z') \qquad \forall z' \in \mathbb{C}^{s+1}. \tag{57}$$

Substituting our expression for Q, we obtain the identity

$$\begin{aligned}\phi(z') &= \int_\sigma d\sigma(z)\,[\phi(z'), :\phi(z)^*\,\phi(z):] \\ &= \int_\sigma d\sigma(z)\,[\phi(z'), \phi(z)^*]\,\phi(z) \\ &\equiv \int_\sigma d\sigma(z)\,K(z',\bar z)\,\phi(z),\end{aligned} \tag{58}$$

where, by the canonical commutation relations,

$$\begin{aligned}&K(z',\bar z) \equiv [\phi(z'),\phi(z)^*] = [\phi(z'),\phi^*(\bar z)] \\ &\quad= (2\pi)^{-2s-2}\int dp'\int dp\,\theta^{-iz'p'}\,\theta^{i\bar z p}\,[\tilde\phi(p'),\tilde\phi(p)^*] \\ &\quad= (2\pi)^{-s}\int dp'\int dp\,\theta^{-iz'p'}\,\theta^{i\bar z p}\,\mathrm{sign}(p_0)\,\delta(p^2-m^2)\,\delta(p'-p) \\ &\quad= \int_{\Omega_m} d\tilde p\ \mathrm{sign}(p_0)\,\theta(yp)\,\theta(y'p)\,\exp[-i(z'-\bar z)p].\end{aligned} \tag{59}$$

K is a distribution on $\mathbb{C}^{s+1}\times\mathbb{C}^{s+1}$ which is piecewise analytic in $\mathcal{T}\times\mathcal{T}$, with

$$K(z',\bar z) = \begin{cases} -i\Delta^+(z'-\bar z;m), & z',z\in\mathcal{T}_+ \\ i\Delta^-(z'-\bar z;m), & z',z\in\mathcal{T}_- \\ 0, & z'\in\mathcal{T}_+, z\in\mathcal{T}_- \\ 0, & z'\in\mathcal{T}_-, z\in\mathcal{T}_+. \end{cases} \tag{60}$$

The two–point functions $-i\Delta^+$ and $i\Delta^-$ are analytic in $\mathcal{T}_+$ and $\mathcal{T}_-$, respectively, and act as reproducing kernels for the subspaces with charge ε and $-\varepsilon$. Because of the above property, it is reasonable to call $K(z',\bar z)$ a reproducing kernel for the field $\phi(z)$, though this differs somewhat from the standard usage of the term as applied to Hilbert spaces (see chapter 1). Note that K propagates positive–frequency components of the field into the forward ("future") tube and negative–frequency components into the backward ("past") tube. This is somewhat reminiscent of the Feynman propagator, but K is a solution of the *homogeneous* Klein–Gordon equation in the real spacetime variables rather than a Green function.

The energy–momentum and angular momentum operators may be likewise expressed as conserved phase–space integrals of the extended field:

$$
\begin{aligned}
P_\mu &= i \int_\sigma d\sigma \ : \phi^* \partial_\mu \phi : \\
M_{\mu\nu} &= i \int_\sigma d\sigma \ : \phi^* (x_\mu \partial_\nu - x_\nu \partial_\mu) \phi : .
\end{aligned}
\tag{61}
$$

Like Q, these may be displayed as regularizations of the usual, more complicated expressions in real spacetime. Note first that P_μ can be re–written as

$$
\begin{aligned}
P_\mu &= i\, A_\lambda^{-1} \int_S \widehat{dx}^{\,\tau} \int_{\Omega_\lambda} \widehat{dy}_\tau \, : \phi^* \, \partial_\mu \phi : \\
&= \frac{i}{2}\, A_\lambda^{-1} \int_S \widehat{dx}^{\,\tau} \int_{\Omega_\lambda} \widehat{dy}_\tau \, : \phi^* \, \partial_\mu \phi - \bar{\partial}_\mu \phi^* \cdot \phi : \\
&= -\frac{1}{2}\, A_\lambda^{-1} \int_S \widehat{dx}^{\,\tau} \int_{\Omega_\lambda} \widehat{dy}_\tau \, \frac{\partial}{\partial y^\mu} \, : \phi^* \, \phi : .
\end{aligned}
\tag{62}
$$

The angular momentum can be recast similarly as

$$
\begin{aligned}
M_{\mu\nu} &= i\, A_\lambda^{-1} \int_S \widehat{dx}^{\,\tau} \int_{\Omega_\lambda} \widehat{dy}_\tau \, : \phi^* \, (x_\mu \partial_\nu - x_\nu \partial_\nu) \, \phi : \\
&= \frac{i}{2}\, A_\lambda^{-1} \int_S \widehat{dx}^{\,\tau} \int_{\Omega_\lambda} \widehat{dy}_\tau \, \left[x_\mu (\partial_\nu - \bar{\partial}_\nu) - x_\nu (\partial_\mu - \bar{\partial}_\mu) \right] : \phi^* \phi : \\
&= -\frac{1}{2}\, A_\lambda^{-1} \int_S \widehat{dx}^{\,\tau} \int_{\Omega_\lambda} \widehat{dy}_\tau \, \left[x_\mu \frac{\partial}{\partial y^\nu} - x_\nu \frac{\partial}{\partial y^\mu} \right] : \phi^* \phi : .
\end{aligned}
\tag{63}
$$

Using $\Omega_\lambda = -\partial B_\lambda$ and applying Stokes' theorem, we therefore have

$$
\begin{aligned}
P_\mu &= \frac{1}{2}\, A_\lambda^{-1} \int_S \widehat{dx}^{\,\tau} \int_{B_\lambda} dy \, \frac{\partial^2}{\partial y^\mu \partial y^\tau} \, : \phi^* \, \phi : \\
&= \int_S \widehat{dx}^{\,\tau} \, T_{\mu\tau}^{(\lambda)}(x),
\end{aligned}
\tag{64}
$$

where

$$
T_{\mu\tau}^{(\lambda)}(x) \equiv \frac{1}{2}\, A_\lambda^{-1} \int_{B_\lambda} dy \, \frac{\partial^2}{\partial y^\mu \partial y^\tau} \, : \phi^* \, \phi : \tag{65}
$$

is a *regularized energy–momentum density tensor* which, incidentally, is automatically symmetric. Similarly,

$$\begin{aligned} M_{\mu\nu} &= \frac{1}{2}\, A_\lambda^{-1} \int_S \widehat{dx}^{\,\tau} \int_{B_\lambda} dy \, \left[x_\mu \frac{\partial^2}{\partial y^\nu \partial y^\tau} - x_\nu \frac{\partial^2}{\partial y^\mu \partial y^\tau} \right] : \phi^* \phi : \\ &= \int_S \widehat{dx}^{\,\tau}\, \Theta^{(\lambda)}_{\mu\nu\tau}(x), \end{aligned} \tag{66}$$

where

$$\begin{aligned} \Theta^{(\lambda)}_{\mu\nu\tau}(x) &\equiv \frac{1}{2}\, A_\lambda^{-1} \int_{B_\lambda} dy \, \left[x_\mu \frac{\partial^2}{\partial y^\nu \partial y^\tau} - x_\nu \frac{\partial^2}{\partial y^\mu \partial y^\tau} \right] : \phi^* \phi : \\ &= x_\mu T^{(\lambda)}_{\nu\tau}(x) - x_\nu T^{(\lambda)}_{\mu\tau}(x) \end{aligned} \tag{67}$$

is a regularized angular momentum density tensor.

5.5. Free Dirac Fields

For simplicity, we specialize in this section (only) to the physical case of three spatial dimensions, $s = 3$. The proper Lorentz group $\mathcal{L}_0$ is then $SO(3,1)_+$, where the plus sign indicates that $\Lambda^0_0 > 0$, so that Λ preserves the orientations of space and time separately. Its universal covering group can be identified with $SL(2,\mathbb{C})$ as follows (Streater and Wightman [1964]): An event $x \in \mathbb{R}^4$ is identified with the Hermitian 2×2 matrix

$$X = x^\mu \sigma_\mu = \begin{pmatrix} x^0 + x^3 & x^1 - ix^2 \\ x^1 + ix^2 & x^0 - x^3 \end{pmatrix} \tag{1}$$

where $\sigma_0 = I$ (2×2 identity) and σ_k ($k = 1, 2, 3$) are the Pauli spin matrices. Note that $\det X = x^2 \equiv x \cdot x$. The action of $SL(2,\mathbb{C})$ on Hermitian 2×2 matrices given by

$$X' = AXA^*, \qquad A \in SL(2,\mathbb{C}), \tag{2}$$

induces a linear transformation on $\mathbb{R}^4$ which we denote by $\pi(A)$:

$$x' = \pi(A)\, x \tag{3}$$

From

$$x'^2 = \det X' = |\det A|^2 \det X = \det X = x^2 \tag{4}$$

it follows that $\pi(A)$ is a Lorentz transformation, and it can easily be seen to be proper. Hence π defines a map

$$\pi\colon SL(2,\mathbb{C}) \to \mathcal{L}_0, \tag{5}$$

which is readily seen to be a group homomorphism. Clearly, $\pi(-A) = \pi(A)$, and it can be shown that if $\pi(A) = \pi(B)$, then $A = \pm B$. Since $SL(2,\mathbb{C})$ is simply connected, it follows that $SL(2,\mathbb{C})$ is the *universal covering group* of $\mathcal{L}_0$, the correspondence being two–to–one.

The relativistic transformation law as stated in section 5.3,

$$\phi(\Lambda x + a) = U(a,\Lambda)\,\phi(x)\,U(a,\Lambda)^*, \tag{6}$$

applies to scalar fields, i.e. fields without any intrinsic orientation or *spin*. To generalize it to fields with spin, note first of all that since the representing operator $U(a,\Lambda)$ occurs quadratically, the law is invariant under $U \to -U$. This means that U could, in fact, be a representation, not of $\mathcal{P}_0$, but of the inhomogeneous version of $SL(2,\mathbb{C})$,

$$\mathcal{P}_2 \equiv \mathbb{R}^4 \circledS\, SL(2,\mathbb{C}), \tag{7}$$

which acts on $\mathbb{R}^4$ by

$$(a,A)\,x = \pi(A)\,x + a. \tag{8}$$

$\mathcal{P}_2$ is the two–fold universal covering group of $\mathcal{P}_0$. A field $\psi(x)$ of arbitrary spin is a distribution taking its values in the tensor product $L(\mathcal{H}) \otimes \mathcal{V}$ of the operator algebra of the quantum Hilbert space $\mathcal{H}$ with some finite–dimensional representation space $\mathcal{V}$ of $SL(2,\mathbb{C})$. The transformation law is

$$U(a,A)\,\psi(x)\,U(a,A)^* = S(A^{-1})\,\psi(\pi(A)x + a), \tag{9}$$

where S is a given representation of $SL(2,\mathbb{C})$ in $\mathcal{V}$. S determines the spin of the field, which can take the values $j = 0, 1/2, 1, 3/2, 2, \ldots$.

The locality condition for the scalar field (axiom 4) can be extended to non–scalar fields as

$$[\psi_\alpha(x), \psi_\beta(x')] = 0 \quad \text{if } (x-x')^2 < 0 \tag{10}$$

where ψ_α are the components of ψ. Now it follows from the axioms that if the field has half–integral spin ($j = 1/2, 3/2, \ldots$), then the above locality condition implies that it is trivial, i.e. that $\psi(x) \equiv 0$. A non–trivial field of half–integral spin can be obtained, however, if we modify the locality condition by replacing commutators with anticommutators:

$$\begin{aligned} \{\psi_\alpha(x), \psi_\beta(x')\} &\equiv \psi_\alpha(x)\,\psi_\beta(x') + \psi_\beta(x')\,\psi_\alpha(x) \\ &= 0 \quad \text{if } (x - x')^2 < 0. \end{aligned} \tag{11}$$

Replacing the commutators with anticommutators means that changing the order in which $\psi(x)$ and $\psi(x')$ are applied to a state vector in Hilbert space merely changes the sign of the vector, which has no observable effect. Hence the physical interpretation that events at spacelike separations cannot influence one another is still valid.

Similarly, for fields of integral spin ($j = 0, 1, \ldots$), the locality condition with anticommutators gives a trivial theory, whereas a non–trivial theory can exist using commutators.

The choice of commutators or anticommutators in the locality condition does, however, have an important physical consequence. For we have seen that the free asymptotic fields can be written as sums of creation and destruction operators for particles and antiparticles. If $x_1, x_2, \cdots x_n$ are n distinct points in the hyperplane $x^0 = 0$ and $\psi_+(x)$ denotes the positive–frequency part of the field (which can be obtained from $\psi(x - iy)$ by taking $y \to 0$ in V'_+), then

$$\psi_+(x_1)^* \psi_+(x_2)^* \cdots \psi_+(x_n)^*\, \Psi_0 \tag{12}$$

is a state with n particles located at these points. Since any two of these points are separated by a spacelike distance, the locality condition implies that this state is *symmetric* with respect to the exchange of any two particles if commutators are used and *antisymmetric* with respect to the exchange if anticommutators are used. Particles whose states are symmetric under exchange are called *Bosons*, and ones antisymmetric under exchange are called *Fermions*. The choice of symmetry or antisymmetry crucially affects the large–scale statistical behavior of the particles. For example, no two Fermions can occupy the same state due to the antisymmetry under exchange; this is the *Pauli exclusion principle*. Hence the choice of commutators or anticommutators is known as the choice of *statistics*, and the above theorem correlating this choice with the spin is known as the *Spin–Statistics theorem* of quantum field theory (Streater and Wightman [1964]).

This theorem is fully supported by experiment, and represents one of the successes of the theory.

The Dirac field is a quantized field with spin 1/2 whose associated particles and antiparticles are typically taken to be electrons and positrons, though it is also used (albeit less accurately) to model neutrons and protons. Our treatment follows the notation used in Itzykson and Zuber [1980], with minor modifications. The free Dirac field is a solution of the Dirac equation

$$(i\partial\!\!\!/ - m)\psi(x) = 0, \tag{13}$$

where

$$\partial\!\!\!/ \equiv \gamma^\mu \frac{\partial}{\partial x^\mu} \tag{14}$$

is the Dirac operator and the γ's are a set of 4×4 Dirac matrices, meaning they satisfy the Clifford condition with respect to the Minkowski metric:

$$\{\gamma^\mu, \gamma^\nu\} \equiv \gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2g^{\mu\nu}. \tag{15}$$

The components of ψ satisfy the Klein–Gordon equation, since $\partial\!\!\!/^{\,2} = \Box$, and the solutions can be written as

$$\psi(x) = \int_{\Omega_m^+} d\tilde{p}\ \left[e^{-ixp}\, u^\alpha(p)\, b_\alpha(p) + e^{ixp}\, v^\alpha(p)\, d^*_\alpha(p)\right], \tag{16}$$

where u^α and v^α are positive– and negative–frequency four–component spinors and summation over the polarization index $\alpha = 1, 2$ is implied. b_α and d_α are operators satisfying the "canonical anticommutation relations"

$$\begin{aligned} &\{b_\alpha(p), b_\beta(q)\} = \{d_\alpha(p), d_\beta(q)\} = \{b_\alpha(p), d_\beta(q)\} = 0, \\ &\{b_\alpha(p), b^*_\beta(q)\} = \{d_\alpha(p), d^*_\beta(q)\} = 2\omega(2\pi)^3\, \delta(\mathbf{p} - \mathbf{q}) \end{aligned} \tag{17}$$

for all $p, q \in \Omega_m^+$. The spinors satisfy the orthogonality and completeness relations

$$\begin{aligned} \bar{u}^\alpha(p)\, u^\beta(p) &= -\bar{v}^\alpha(p)\, v^\beta(p) = \delta^{\alpha\beta} \\ u^\alpha(p) \otimes \bar{u}^\alpha(p) &= \frac{\not{p} + m}{2m} \\ v^\alpha(p) \otimes \bar{v}^\alpha(p) &= \frac{\not{p} - m}{2m}, \end{aligned} \tag{18}$$

where a summation on α is implied in the last two equations and the adjoint spinors are defined by

$$\bar{u}^\alpha(p) = u^\alpha(p)^* \gamma^0, \quad \bar{v}^\alpha(p) = v^\alpha(p)^* \gamma^0. \tag{19}$$

In addition,

$$\bar{u}^\alpha(p)\, \gamma_\mu\, u^\beta(p) = \bar{v}^\alpha(p)\, \gamma_\mu\, v^\beta(p) = \frac{p_\mu}{m}\, \delta^{\alpha\beta}. \tag{20}$$

$b_\alpha(p)$ and $d_\alpha(p)$ are interpreted as annihilation operators for particles and antiparticles, respectively, while their adjoints are creation operators. The adjoint field is defined by

$$\overline{\psi(x)} = \psi(x)^*\, \gamma^0 \tag{21}$$

and satisfies

$$i \frac{\partial \bar{\psi}}{\partial x^\mu}\, \gamma^\mu = m \bar{\psi}. \tag{22}$$

The particle– and antiparticle number operators are now

$$\begin{aligned} N_+ &= \int_{\Omega_m^+} d\tilde{p}\; b_\alpha^*(p)\, b_\alpha(p) \\ N_- &= \int_{\Omega_m^+} d\tilde{p}\; d_\alpha^*(p)\, d_\alpha(p), \end{aligned} \tag{23}$$

and the charge operator is

$$Q = \varepsilon(N_+ - N_-). \tag{24}$$

As for the Klein–Gordon field, we wish to give a phase–space representation of Q. The first step is to extend $\psi(x)$ to $\mathbb{C}^4$ using the Analytic–Signal transform, which gives

$$\psi(z) = \int_{\Omega_m^+} d\tilde{p}\ \left[\theta^{-izp}\, u^\alpha(p)\, b_\alpha(p) + \theta^{izp}\, v^\alpha(p)\, d^*_\alpha(p)\right]. \tag{25}$$

Again, the extended field is analytic in $\mathcal{T}$, with the parts in $\mathcal{T}_+$ and $\mathcal{T}_-$ containing only positive and negative frequncies, respectively. Using the above orthogonality relations, as well as

$$\int_{\sigma_+} d\sigma\, e^{i\bar{z}p - izq} = \int_{\sigma_-} d\sigma\, e^{izp - i\bar{z}q} = 2\omega(\mathbf{p})\,(2\pi)^3\,\delta(\mathbf{p}-\mathbf{q}) \tag{26}$$

for $p, q \in \Omega_m^+$, we obtain the following expressions for the particle– and antiparticle number operators as phase–space integrals:

$$\begin{aligned} N_+ &= \int_{\sigma_+} d\sigma\, \overline{\psi(z)}\,\psi(z) \equiv \int_{\sigma_+} d\sigma\ :\overline{\psi(z)}\,\psi(z): \\ N_- &= \int_{\sigma_-} d\sigma\ :\overline{\psi(z)}\,\psi(z):, \end{aligned} \tag{27}$$

where the fields in the first integral are already in normal order and the second integral involves two changes of sign: one due to the normal ordering, and another due to the orthogonality relation for the v^α's. The charge operator can therefore be given the following compact expression as a phase–space integral over the oriented phase space $\sigma = \sigma_+ - \sigma_-$:

$$Q = \varepsilon \int_\sigma d\sigma\ :\bar{\psi}\,\psi: = \varepsilon \int_\sigma d\sigma\, \rho(z), \tag{28}$$

where $\rho \equiv :\bar{\psi}\psi:$ is the *scalar phase–space charge density*. The usual expression for the charge as an integral over a configuration space S is

$$Q_{\text{usual}} = \varepsilon \int_S \widehat{dx}_\mu :\overline{\psi(x)}\,\gamma^\mu\,\psi(x): \equiv \int_S \widehat{dx}_\mu J^\mu(x). \tag{29}$$

To compare these two expressions, we again use $\Omega_\lambda = -\partial B_\lambda$ and invoke Stokes' theorem:

$$\begin{aligned} Q &= \varepsilon A_\lambda^{-1} \int_S \widehat{dx}_\mu \int_{\Omega_\lambda} \widehat{dy}^{\,\mu} : \bar\psi\,\psi : \\ &= -\varepsilon A_\lambda^{-1} \int_S \widehat{dx}_\mu \int_{B_\lambda} dy\, \frac{\partial}{\partial y_\mu} : \bar\psi\,\psi : . \end{aligned} \tag{30}$$

Define the *phase–space current density*

$$j^\mu(z) \equiv 2m\varepsilon : \overline{\psi(z)}\gamma^\mu\,\psi(z) :, \tag{31}$$

where the factor $2m$ is included to give j^μ the correct physical dimensions, given our normalization. Note that $j^\mu(z)$ is conserved in spacetime, i.e.

$$\begin{aligned} \frac{\partial j^\mu}{\partial x^\mu} &= (\bar\partial_\mu + \partial_\mu) j^\mu \\ &= 2m\varepsilon : \frac{\partial\bar\psi}{\partial\bar z^\mu}\,\gamma^\mu\,\psi + \bar\psi\gamma^\mu \frac{\partial\psi}{\partial z^\mu} : \\ &= 0 \end{aligned} \tag{32}$$

by the Dirac equation combined with the analyticity of ψ in $\mathcal{T}$. The same combination also implies

$$\begin{aligned} \frac{1}{2}\, j^\mu(z) &= \varepsilon : \bar\psi\gamma^\mu i\gamma^\nu\partial_\nu\psi : \\ &= i\varepsilon\partial_\nu : \bar\psi\gamma^\mu\gamma^\nu\psi : \\ &= i\varepsilon\partial_\nu : \bar\psi(g^{\mu\nu} - i\sigma^{\mu\nu})\psi : \\ &= i\varepsilon\partial^\mu : \bar\psi\psi : + \varepsilon\partial_\nu : \bar\psi\sigma^{\mu\nu}\psi :, \end{aligned} \tag{33}$$

where

$$\sigma^{\mu\nu} = \frac{i}{2}\,[\gamma^\mu, \gamma^\nu] \tag{34}$$

are the spin matrices. The real part of this equation gives a phase–space version of the *Gordon identity*

$$\begin{aligned} j^\mu(z) &= i\varepsilon(\partial^\mu - \bar\partial^\mu) : \bar\psi\psi : + \varepsilon(\partial_\nu + \bar\partial_\nu) : \bar\psi\sigma^{\mu\nu}\psi : \\ &= -\varepsilon\frac{\partial\rho}{\partial y_\mu} + \varepsilon\frac{\partial}{\partial x^\nu} : \bar\psi\sigma^{\mu\nu}\psi : . \end{aligned} \tag{35}$$

The two terms are conserved separately, since

$$\begin{aligned} \frac{\partial^2 \rho}{\partial x^\mu \partial y_\mu} &= i\varepsilon(\Box_{\bar{z}} - \Box_z) :\bar{\psi}\psi: = 0 \\ \frac{\partial^2}{\partial x^\mu \partial x^\nu} &: \bar{\psi}\sigma^{\mu\nu}\psi : = 0, \end{aligned} \tag{36}$$

and the second term, which is due to spin, does not contribute to the total charge since it is a pure divergence with respect to x. Thus

$$Q = A_\lambda^{-1} \int_S \widehat{dx}_\mu \int_{B_\lambda} dy\, j^\mu(z) = \int_S \widehat{dx}_\mu\, J^\mu_{(\lambda)}(x), \tag{37}$$

where

$$J^\mu_{(\lambda)}(x) \equiv A_\lambda^{-1} \int_{B_\lambda} dy\ j^\mu(x - iy) \tag{38}$$

is a "regularized" spacetime current.

Note: The Dirac equation can also be written in the manifestly gauge–invariant form

$$i\partial_\mu j^\mu(z) = 2m\varepsilon : \bar{\psi}\gamma^\mu \partial_\mu \psi : = 2m^2 \rho(z).\ \# \tag{39}$$

Again, ψ is a "root vector" of the charge operator, since it removes a charge ε from any state to which it is applied:

$$[\psi(z'), Q] = \varepsilon\psi(z') \qquad \forall z' \in \mathbb{C}^4. \tag{40}$$

Substituting for Q the above phase–space integral and using the commutator identity

$$[A, BC] = \{A, B\}C - B\{A, C\} \tag{41}$$

and the canonical anticommutation relations, we obtain

$$\psi(z') = \int_\sigma d\sigma\, \{\psi(z'), \overline{\psi(z)}\}\, \psi(z) \equiv \int_\sigma d\sigma\, K_D(z', \bar{z})\psi(z), \tag{42}$$

where the "reproducing kernel" for the Dirac field is a matrix–valued distribution on $\mathbb{C}^4 \times \mathbb{C}^4$ given by

$$\begin{aligned} K_D(z',\bar{z}) &= \{\psi(z'), \overline{\psi(z)}\} \\ &= \int_{\Omega_m^+} d\tilde{p}\, \Big[\theta(y'p)\,\theta(yp)\, e^{-i(z'-\bar{z})p}\, u^\alpha \otimes \bar{u}^\alpha \\ &\qquad + \theta(-y'p)\,\theta(-yp)\, e^{i(z'-\bar{z})p}\, v^\alpha \otimes \bar{v}^\alpha\Big] \\ &= \left(\frac{i\partial\!\!\!/' + m}{2m}\right) K(z',\bar{z}). \end{aligned} \tag{43}$$

Here, K is the reproducing kernel for the Klein–Gordon field and $\partial\!\!\!/'$ is the Dirac operator with respect to the real part x' of z'. Like K, K_D is piecewise holomorphic in $z' - \bar{z}$ for $z', z \in \mathcal{T}$. Another form of the reproducing relation can be obtained by substituting the more complicated expression for Q given by eq. (37) into eq. (40):

$$\psi(z') = 2mA_\lambda^{-1} \int_S \widehat{dx}_\mu \int_{B_\lambda} dy\; K_D(z',\bar{z})\, \gamma^\mu\, \psi(z). \tag{44}$$

This form is closer to the usual relation.

The energy–momentum and angular momentum operators for the Dirac field can likewise be represented by phase–space integrals as

$$\begin{aligned} P_\mu &= \int_\sigma d\sigma\; :\bar{\psi}\, i\partial_\mu \psi: \\ M_{\mu\nu} &= \int_\sigma d\sigma\; :\bar{\psi}\,(ix_\mu\partial_\nu - ix_\nu\partial_\mu + \tfrac{1}{2}\sigma_{\mu\nu})\,\psi: . \end{aligned} \tag{45}$$

More generally, let $\psi(z)$ represent either a Klein–Gordon field (in which case $\bar{\psi}$ will mean ψ^*) or a Dirac field, and let T_a be the local generators of an arbitrary *internal or external* symmetry group, so that the infinitesimal change in $\psi(z)$ is given by

$$\delta\psi(z) = -i\epsilon^a T_a \psi(z). \tag{46}$$

For example, T_a is multiplication by ε for $U(1)$ gauge symmetry, $T_\mu = i\partial_\mu$ for spacetime translations (where the derivative is with respect to x^μ), etc. (In case the theory has an internal symmetry higher than $U(1)$, of course, ψ must have extra indices since it must be valued in a representation space of the corresponding Lie algebra.) The generators satisfy the Lie relations

$$[T_a, T_b] = C^c_{ab}\, T_c, \tag{47}$$

where C^c_{ab} are the structure constants. Then we claim that the conserved *global* field observable corresponding to T_a is

$$Q_a = \int_\sigma d\sigma\ : \bar{\psi}\, T_a \psi : . \tag{48}$$

For this implies

$$[\psi(z'), Q_a] = \int_\sigma d\sigma\, K_D(z', \bar{z})\, T_a \psi(z), \tag{49}$$

where K_D is replaced by K if ψ is a Klein–Gordon field. Since T_a generates a symmetry, it follows that $T_a\psi(z)$ is also a solution of the appropriate wave equation, hence it is reproduced by K_D:

$$\int_\sigma d\sigma\, K_D(z', \bar{z}) T_a \psi(z) = T_a \psi(z'). \tag{50}$$

Therefore Q_a has the required property

$$[\psi(z'), Q_a] = T_a \psi(z'). \tag{51}$$

It can furthermore be checked that

$$[Q_a, Q_b] = \int_\sigma d\sigma\ : \bar{\psi}\, [T_a, T_b]\, \psi := C^c_{ab}\, Q_c, \tag{52}$$

hence the mapping $T_a \mapsto Q_a$ is a Lie algebra homomorphism.

Finally, we show that due to the separation of positive and negative frequencies in $\mathcal{T}$, the interference effect known as Zitterbewegung does not occur for Fermions in the phase–space formalism. Let S_t be the configuration space defined by $x^0 = t$. Then the components of the "regularized" three–current at time t are

$$J^k_{(\lambda)}(t) = 2mA_\lambda^{-1} \int_{S_t} d^3\mathbf{x} \int_{B_\lambda} dy : \bar{\psi}\gamma^k \psi :, \tag{53}$$

and a straightforward computation gives

$$J^k_{(\lambda)}(t) = \int_{\Omega^+_m} d\tilde{p}\ \left(\frac{p^k}{m}\right) \left[b^*_\alpha(p)\, b_\alpha(p) - d^*_\alpha(p)\, d_\alpha(p)\right]. \tag{54}$$

The right–hand side is independent of t, hence no Zitterbewegung occurs. In *real* spacetime, Zitterbewegung is the result of the inevitable interference between the positive– and negative–frequency components of ψ. Its absence in complex spacetime is due to the *polarization* of the positive and negative frequencies of ψ into $\mathcal{T}_+$ and $\mathcal{T}_-$, respectively.

In the usual theory, Zitterbewegung is shown to occur in the single–particle theory; the above computation can be repeated for the classical (i.e., "first–quantized") Dirac field, with an identical result except for a change in sign in the second term due to the commutation of d^*_α and d_α. Alternatively, the above argument also implies the absence of Zitterbewegung for the one–particle and one–antiparticle states of the Dirac field.

5.6. Interpolating Particle Coherent States

We now return to the interpolating charged scalar field ϕ. The asymptotic fields satisfy the Klein–Gordon equation,

$$(\Box + m^2)\,\phi_{\rm in}(x) = 0, \quad (\Box + m^2)\,\phi_{\rm out}(x) = 0 \tag{1}$$

and have the same vacuum expectation values as the free Klein–Gordon field discussed in section 5.4. Hence, by Wightman's reconstruction theorem (Streater and Wightman [1964]), these three fields are unitarily related. We identify the free field of section 5.4 with $\phi_{\rm in}$. Then there is a unitary operator S such that

$$\phi_{\rm out}(x) = S\,\phi_{\rm in}(x)\,S^*. \tag{2}$$

S is known as the *scattering operator.*

Define the *source field* $j(x)$ by

$$j(x) \equiv (\Box + m^2)\,\phi(x). \tag{3}$$

It is a measure of the extent of the interaction at x, and by axiom 5,

$$j(x) \to 0 \quad \text{(weakly) as } x^0 \to \pm\infty. \tag{4}$$

Note that we are not making any additional assumptions about j. If j is a known *function* (i.e., if it is a multiple of the identity on $\mathcal{H}$ for each x), then it acts as an *external source* for ϕ. If, on the other hand, j is a local function of ϕ such as $:\phi^3:$, it represents a self–interaction of

ϕ. In any case, the above equations can be "solved" using the Green functions of the Klein–Gordon operator, which satisfy

$$(\Box_x + m^2)\, G(x) = \delta(x). \tag{5}$$

In general, we have formally

$$\phi(x) = \phi_0(x) + \int dx'\, G(x - x')\, j(x'), \tag{6}$$

where ϕ_0 is a free field determined by the initial or boundary conditions at infinity used to determine G. The *retarded* Green function (we are back to s spatial dimensions) is defined as

$$G_{\rm ret}(x) = (2\pi)^{-s-1} \int_{\mathbb{R}^{s+1}} dp\, \frac{e^{-ixp}}{(p_+^2 - m^2)}, \tag{7}$$

where

$$p_+ \equiv (p_0 + i\epsilon, \mathbf{p}) \tag{8}$$

with $\epsilon > 0$ and the limit $\epsilon \downarrow 0$ is taken after the integral is evaluated. $G_{\rm ret}$ propagates both positive and negative frequencies *forward* in time, which means that it is *causal*, i.e. vanishes when $x_0 < 0$. Since it is also Lorentz–invariant, it follows that

$$G_{\rm ret}(x - x') = 0 \quad \text{unless } x - x' \in \overline{V_+'}. \tag{9}$$

$G_{\rm ret}(x - x')$ is interpreted as the causal effect at x due to a unit disturbance at x'. The corresponding choice of free field ϕ_0 is $\phi_{\rm in}$, hence

$$\phi(x) = \phi_{\rm in}(x) + \int dx'\, G_{\rm ret}(x - x')\, j(x'). \tag{10}$$

If j is a known external source, this gives a complete solution for $\phi(x)$. If j is a known function of ϕ, it merely gives an integral equation which ϕ must satisfy.

Similarly, the *advanced* Green function is defined by

$$G_{\rm adv}(x - x') = (2\pi)^{-s-1} \int_{\mathbb{R}^{s+1}} dp\, \frac{e^{-i(x-x')p}}{(p_-^2 - m^2)}, \tag{11}$$

with $p_- \equiv (p_0 - i\epsilon, \mathbf{p})$ and $\epsilon \downarrow 0$, and propagates both positive and negative frequencies *backward* in time, which means it is *anticausal.* The corresponding free field is $\phi_{\rm out}$, hence

$$\phi(x) = \phi_{\rm out}(x) + \int dx'\, G_{\rm adv}(x - x')\, j(x'). \tag{12}$$

Let us now apply the Analytic–Signal transform to both of these equations:

$$\begin{aligned} \phi(z) &= \phi_{\rm in}(z) + \int dx'\, G_{\rm ret}(z - x')\, j(x') \\ \phi(z) &= \phi_{\rm out}(z) + \int dx'\, G_{\rm adv}(z - x')\, j(x'), \end{aligned} \tag{13}$$

where (with $z = x - iy$)

$$\begin{aligned} G_{\rm ret}(z - x') &\equiv \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{d\tau}{\tau - i}\, G_{\rm ret}(x - \tau y - x') \\ &= (2\pi)^{-s-1} \int_{\mathbb{R}^{s+1}} dp\, \frac{\theta(yp)\, e^{-i(z-x')p}}{(p_+^2 - m^2)} \end{aligned} \tag{14}$$

and

$$\begin{aligned} G_{\rm adv}(z - x') &\equiv \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{d\tau}{\tau - i}\, G_{\rm adv}(x - \tau y - x') \\ &= (2\pi)^{-s-1} \int_{\mathbb{R}^{s+1}} dp\, \frac{\theta(yp)\, e^{-i(z-x')p}}{(p_-^2 - m^2)}. \end{aligned} \tag{15}$$

Since the Analytic–Signal transform involves an integration over the entire line $x(\tau) = x - \tau y$, the effect of $G_{\rm ret}(z - x')$ is no longer causal when regarded as a function of z and x'. Rather, it might be interpreted as the causal effect of a unit disturbance at x' on the line parametrized by z. (Note that only those values of τ for which $x - \tau y - z' \in \overline{V_+'}$ contribute to the integral.) A similar statement goes for $G_{\rm adv}(z - x')$.

Whereas $\phi_{\rm in}(z)$ and $\phi_{\rm out}(z)$ are holomorphic in $\mathcal{T}$, $\phi(z)$ is *not* (unless $j(x) \equiv 0$), since $G_{\rm ret}(z - x')$ and $G_{\rm adv}(z - x')$ are not holomorphic. This breakdown of holomorphy in the presence of interactions is by

now expected. Of course ϕ, $G_{\rm ret}$ and $G_{\rm adv}$ are all holomorphic along the vector field y, as are all Analytic–Signal transforms.

In Wightman field theory, the vacua $\Psi_0^{\rm in}$, $\Psi_0^{\rm out}$ and Ψ_0 of the in–, out– and interpolating fields all coincide (the theory is "already–renormalized"). Let us define the *asymptotic particle coherent states* by

$$\begin{aligned} e^+_{{\rm in},z} &= \phi_{\rm in}(z)^*\Psi_0 \\ e^-_{{\rm in},z} &= \phi_{\rm in}(z)\Psi_0 \\ e^+_{{\rm out},z} &= \phi_{\rm out}(z)^*\Psi_0 \\ e^-_{{\rm out},z} &= \phi_{\rm out}(z)\Psi_0. \end{aligned} \tag{16}$$

We will refer to

$$e^+_z = \phi(z)^*\Psi_0, \quad e^-_z = \phi(z)\Psi_0 \tag{17}$$

as the *interpolating particle coherent states*. By eq. (13),

$$\begin{aligned} e^+_z &= e^+_{{\rm in},z} + \int dx'\, \overline{G_{\rm ret}(z-x')}\, j(x')^*\, \Psi_0 \\ &= e^+_{{\rm out},z} + \int dx'\, \overline{G_{\rm adv}(z-x')}\, j(x')^*\, \Psi_0 \end{aligned} \tag{18}$$

and

$$\begin{aligned} e^-_z &= e^-_{{\rm in},z} + \int dx'\, G_{\rm ret}(z-x')\, j(x')\, \Psi_0 \\ &= e^-_{{\rm out},z} + \int dx'\, G_{\rm adv}(z-x')\, j(x')\, \Psi_0. \end{aligned} \tag{19}$$

From the definitions it follows that

$$\begin{aligned} \overline{G_{\rm ret}(z-x')} &= G_{\rm adv}(x'-\bar z) \\ \overline{G_{\rm adv}(z-x')} &= G_{\rm ret}(x'-\bar z), \end{aligned} \tag{20}$$

hence eq. (18) can be rewritten as

$$\begin{aligned} e^+_z &= e^+_{{\rm in},z} + \int dx'\, G_{\rm adv}(x'-\bar z)\, j(x')^*\, \Psi_0 \\ &= e^+_{{\rm out},z} + \int dx'\, G_{\rm ret}(x'-\bar z)\, j(x')^*\, \Psi_0. \end{aligned} \tag{21}$$

Eqs. (19) and (21) display the interpolating character of $e_z^{\pm}$. Note that when $j(x)$ is an external source, then the interpolating particle coherent states differ from the asymptotic ones by a multiple of the vacuum.

As in the case of the free theory, a general state with a single positive charge ε can be written in the form

$$\Psi_f^+ = \phi^*(f)\,\Psi_0. \tag{22}$$

For interacting fields, this may, in general, no longer be interpreted as a one–particle state, since no particle–number operator exists.* But the charge operator does exist since charge (unlike particle–number) is conserved in general, due to gauge invariance; hence Ψ_f^+ makes sense as an eigenvector of charge with eigenvalue ε. Ψ_f^+ can be expressed in terms of particle coherent states as

$$\begin{aligned}\tilde{f}(z) &\equiv \langle\, e_z^+ \,|\, \Psi_f^+ \,\rangle \\ &= \langle\, e_{\mathrm{in},z}^+ \,|\, \Psi_f^+ \,\rangle + \int dx'\, G_{\mathrm{ret}}(z-x')\,\langle\, \Psi_0 \,|\, j(x')\,\Psi_f^+ \,\rangle \\ &= \langle\, e_{\mathrm{out},z}^+ \,|\, \Psi_f^+ \,\rangle + \int dx'\, G_{\mathrm{adv}}(z-x')\,\langle\, \Psi_0 \,|\, j(x')\,\Psi_f^+ \,\rangle.\end{aligned} \tag{23}$$

$\tilde{f}(z)$ satisfies the inhomogeneous equations

$$\begin{aligned}(\Box_x + m^2)\,\tilde{f}(z) &= \int dx'\,(\Box_x + m^2)\,G_{\mathrm{ret}}(z-x')\,\langle\,\Psi_0\,|\,j(x')\,\Psi_f^+\,\rangle \\ &= \int dx'\,(\Box_x + m^2)\,G_{\mathrm{adv}}(z-x')\,\langle\,\Psi_0\,|\,j(x')\,\Psi_f^+\,\rangle.\end{aligned} \tag{24}$$

But from the definitions it follows that

* If the spectrum Σ contains an isolated mass shell Ω_m^+ and $\tilde{f}(p)$ is concentrated around Ω_m^+, then Ψ_f^+ is, in fact, a one–particle state. This is the starting point of the Haag–Ruelle scattering theory (Jost [1965]). I thank R. F. Streater for this remark.

$$\begin{aligned}(\Box_x + m^2)\, G_{\rm ret}(z-x') &= (\Box_x + m^2)\, G_{\rm adv}(z-x') \\ &= (2\pi)^{-s-1} \int_{\mathbb{R}^{s+1}} dp\ \theta(yp)\, e^{-i(z-x')p} \\ &\equiv \delta(z-x'),\end{aligned} \tag{25}$$

where the last equation is a *definition* of $\delta(z-x')$ as the Analytic–Signal transform with respect to x of $\delta(x-x')$. The above is easily seen to reduce to

$$(\Box_x + m^2)\, \tilde{f}(z) = \langle\, \Psi_0 \,|\, j(z)\, \Psi_f^+ \,\rangle, \tag{26}$$

where $j(z)$ is the Analytic–Signal transform of $j(x)$. Equivalently, eq. (3) can be extended to $\mathbb{C}^{s+1}$ by applying the Analytic–Signal transform, giving

$$(\Box_x + m^2)\, \phi(z) = j(z), \tag{27}$$

hence

$$(\Box_x + m^2)\, \tilde{f}(z) = \langle\, \Psi_0 \,|\, (\Box_x + m^2)\, \phi(z) \,|\, \Psi_f^+ \,\rangle = \langle\, \Psi_0 \,|\, j(z)\, \Psi_f^+ \,\rangle. \tag{28}$$

For a known external source, this is a "perturbed" Klein–Gordon equation for $\tilde{f}(z)$; if j depends on ϕ, it appears to be of little value.

5.7. Field Coherent States and Functional Integrals

So far, all our coherent states have been states with a single particle or antiparticle. In this section, we construct coherent states in which the entire field participates, involving an indefinite number of particles. We do so first for a neutral free Klein–Gordon field (or a generalized free field; see section 5.3), then for a free charged scalar field. A similar construction works for Dirac fields, but the "functions" labeling the coherent states must then anticommute instead of being "classical" functions and a generalized type of functional integral must be used (Berezin [1966], Segal [1956b, 1965]). We also indulge in some speculation on generalizing the construction to interpolating fields.

An extended neutral free Klein–Gordon field satisfies the canonical commutation relations

$$\begin{aligned} [\phi(z), \phi(z')] &= 0 \\ [\phi(z), \phi(z')^*] &= K(z, \bar{z}') = -i\Delta^+(z - \bar{z}') \end{aligned} \tag{1}$$

for all $z, z' \in \mathcal{T}_+$, as well as the reality condition $\phi(z)^* = \phi(\bar{z})$. The basic idea is that since all the operators $\phi(z)$ ($z \in \mathcal{T}_+$) commute, it *may* be possible to find a total set of simultaneous eigenvectors for them. This is not guaranteed, since $\phi(z)$ is not self–adjoint (it is not even *normal*, by eq. (1)) and, in any case, it is unbounded and thus may present us with domain problems. However, this hope is realized by explicitly constructing such eigenvectors. This construction mimics that of the canonical coherent states in section 3.4, which used the lowering and raising operators A and A^*. As in the case of finitely many degrees of freedom, the canonical commutation relations mean that ϕ^* acts as a *generator of translations* in the space in which ϕ is "diagonal." The construction proceeds as follows: Let $\hat{f}(\mathbf{p})$ be a function on $\mathbb{R}^s$, which will also be regarded as a function on Ω_m^+. To simplify the analysis, we assume to begin with that $\hat{f}$ is a (complex–valued) Schwartz test function, although this will be relaxed later. $\hat{f}$ determines a holomorphic positive–energy solution of the Klein–Gordon equation,

$$f(z) = \int_{\Omega_m^+} d\tilde{p}\; e^{-izp}\, \hat{f}(\mathbf{p}). \tag{2}$$

Define

$$\phi^*(f) \equiv \int_{\Omega_m^+} d\tilde{p}\; a^*(p)\, \hat{f}(\mathbf{p}) = \int_{\sigma_+} d\sigma\, \phi(z)^*\, f(z), \tag{3}$$

where σ_+ is any particle phase space and the second equality follows from theorem 4.10 and its corollary. (*Note:* this is *not* the same as the smeared field in real spacetime, since the latter would involve an integration over time, which diverges when f is itself a solution rather than a test function in spacetime.) The canonical commutation relations imply that for $z \in \mathcal{T}_+$,

$$[\phi(z), \phi^*(f)] = \int_{\sigma_+} d\sigma(z') K(z, \bar{z}')\, f(z') = f(z), \tag{4}$$

and for $n \geq 1$,

$$[\phi(z), \phi^*(f)^n] = f(z) \cdot n\phi^*(f)^{n-1}. \tag{5}$$

We now define the *field coherent states* of ϕ by the formal expression

$$E^f = e^{\phi^*(f)}\, \Psi_0. \tag{6}$$

Then if $z \in \mathcal{T}_+$, so that $\phi(z)\, \Psi_0 = 0$, eq. (5) implies that

$$\begin{aligned} \phi(z)\, E^f &= [\phi(z), e^{\phi^*(f)}]\, \Psi_0 \\ &= f(z)\, E^f. \end{aligned} \tag{7}$$

Hence E^f is a common eigenvector of all the operators $\phi(z),\ z \in \mathcal{T}_+$. This eigenvalue equation implies that the *state* corresponding to E^f is left unchanged by the removal of a single particle, which requires that E^f be a superposition of states with $0, 1, 2, \cdots$ particles. Indeed,

$$E^f = \sum_{n=0}^{\infty} \frac{1}{n!}\, \phi^*(f)^n\, \Psi_0. \tag{8}$$

The projection of E^f to the one–particle subspace can be obtained by using the particle coherent states e_z:

$$\begin{aligned} \langle\, e_z \,|\, E^f \,\rangle &= \langle\, \Psi_0 \,|\, \phi(z)\, E^f \,\rangle \\ &= f(z) \langle\, \Psi_0 \,|\, E^f \,\rangle = f(z), \end{aligned} \tag{9}$$

where the last equality follows from $\phi(\bar{f})\, \Psi_0 \equiv (\phi^*(f))^*\, \Psi_0 = 0$. More generally, the n–particle component of E^f is given by projecting to the n–particle coherent state

$$e_{z_1 z_2 \cdots z_n} \equiv \phi(z_1)^* \phi(z_2)^* \cdots \phi(z_n)^*\, \Psi_0, \tag{10}$$

which gives

$$\langle\, e_{z_1 z_2 \cdots z_n} \,|\, E^f \,\rangle = f(z_1) f(z_2) \cdots f(z_n), \tag{11}$$

so all particles are in the same state f and the entire system of particles is *coherent!* Similar states have been found to be very useful in the analysis of the phenomenon of coherence in quantum optics (Glauber [1963], Klauder and Sudarshan [1968]), where the name "coherent states" in fact originated. In the usual treatment, the positive–frequency components have to be separated out "by hand" using their

Fourier representation, since one is dealing with the fields in *real* spacetime. For us, this separation occured automatically though the use of the Analytic–Signal transform, i.e. $\phi^*(f)$ can be defined directly as an integral of $f(z)$ over σ_+. (This would remain true even if f had a negative–frequency component, since the integration over σ_+ would still restrict f to positive frequencies.)

The inner product of two field coherent states can be computed as follows. Note first that if $g(z)$ is another positive–energy solution, then

$$\begin{aligned} \phi(\bar{f})\, E^g &\equiv \int_{\sigma_+} d\sigma\, \overline{f(z)}\, \phi(z)\, E^g \\ &= \int_{\sigma_+} d\sigma\, \overline{f(z)}\, g(z)\, E^g \\ &= \langle\, f \,|\, g \,\rangle\, E^g, \end{aligned} \tag{12}$$

where, by theorem 4.10,

$$\langle\, f \,|\, g \,\rangle \equiv \int_{\sigma_+} d\sigma\, \overline{f(z)}\, g(z) = \int_{\Omega_m^+} d\tilde{p}\; \overline{\hat{f}(\mathbf{p})}\, \hat{g}(\mathbf{p}). \tag{13}$$

Hence

$$\begin{aligned} \langle\, E^f \,|\, E^g \,\rangle &= \langle\, \Psi_0 \,|\, e^{\phi(\bar{f})}\, E^g \,\rangle \\ &= e^{\langle\, f \,|\, g \,\rangle}. \end{aligned} \tag{14}$$

Thus E^f belongs to $\mathcal{H}$ (i.e., is normalizable) if and only if $\hat{f}(\mathbf{p})$ belongs to $L^2_+(d\tilde{p})$ or, equivalently, $f(z)$ belongs to the one–particle space $\mathcal{K}$ of holomorphic positive–energy solutions. If we suppose this to be the case for the time being, then the field coherent states E^f are parametrized by the vectors $\hat{f} \in L^2_+(d\tilde{p})$ or $f \in \mathcal{K}$. Next, we look for a resolution of unity in $\mathcal{H}$ in terms of the E^f's. The standard procedure (section 1.3) would be to look for an appropriate measure $d\mu(f)$ on $\mathcal{K}$. Actually, it turns out that due to the infinite dimensionality of $\mathcal{K}$, a larger space $\mathcal{K}'_0 \supset \mathcal{K}$ will be needed to support $d\mu$. Thus, for the time being, we leave the domain of integration unspecified and write formally

$$\int_{\mathcal{K}} d\mu(f)\; |\, E^f \,\rangle\langle\, E^f \,| \;= I_{\mathcal{H}}, \tag{15}$$

where $d\mu$ is to be found. Taking the matrix element of this equation between the states E^h and E^g, we obtain

$$\int_{\mathcal{K}} d\mu(f)\; e^{\langle\, h\,|\,f\,\rangle + \langle\, f\,|\,g\,\rangle} = e^{\langle\, h\,|\,g\,\rangle}. \tag{16}$$

With $h = -g$ this gives

$$\int_{\mathcal{K}} d\mu(f)\; e^{\langle\, f\,|\,g\,\rangle - \langle\, g\,|\,f\,\rangle} = e^{-\langle\, g\,|\,g\,\rangle} \equiv S[g]. \tag{17}$$

The left–hand side is an infinite–dimensional version of the Fourier transform of $d\mu$, as becomes apparent if we decompose f and g into their real and imaginary parts. The Fourier transform of a measure is called its *characteristic function.* Hence we conclude that a *necessary* condition for the existence of $d\mu$ is that its characteristic function be $S[g]$. In turn, a function must satisfy certain conditions in order to be the characteistic function of a measure. In the finite–dimensional case, Bochner's theorem (Yosida [1971]) guarantees the existence of the measure if these conditions are satisfied. If the infinite–dimensional space of f's is replaced by $\mathbb{C}^n$, the above relation would uniquely determine $d\mu$ as a Gaussian measure. For the identity

$$\int_{\mathbb{C}^n} \det A^{-1} d^{2n}\zeta\, \exp[-\pi(\zeta - A\xi)^*\, A^{-1}(\zeta - A\xi)] = 1, \tag{18}$$

where A is a positive–definite matrix, implies

$$\int_{\mathbb{C}^n} d\mu(\zeta)\, e^{\pi(\zeta^*\xi + \xi^*\zeta)} = e^{\pi\xi^* A\xi}, \tag{19}$$

with

$$d\mu(\zeta) = \det A^{-1}\; \exp[-\pi\zeta^* A^{-1}\zeta]\, d^{2n}\zeta. \tag{20}$$

The integral in eq. (19) is entire in the variables ξ and ξ^* *separately,* hence it can be analytically continued to $\xi^* \to -\xi^*$, giving

$$\int_{\mathbb{C}^n} d\mu(\zeta)\, e^{\pi(\zeta^*\xi - \xi^*\zeta)} = e^{-\pi\xi^* A\xi}. \tag{21}$$

If $\zeta = \alpha + i\beta$ and $\xi = u + iv$ with $\alpha, \beta, u, v \in \mathbb{R}^n$, then

$$\int_{\mathbb{R}^{2n}} d\mu(\alpha,\beta)\, e^{2\pi i(\alpha v-\beta u)} = e^{-\pi(uAu+vAv)} \equiv S(v,-u), \tag{22}$$

showing that $S(v,-u)$ is the Fourier transform of $d\mu(\alpha,\beta)$.

If A is merely positive–semidefinite, i.e., if it is singular, then $d\mu$ still exists but is concentrated on the *range* of A, and A^{-1} makes sense as a map from this range to the orthogonal complement of the kernel of A.

Eq. (19) is a finite–dimensional version of eq. (16) (with $g = h$). In going to the infinite–dimensional case, two separate complications arise. First of all, recall that in finite dimesions, the Fourier transform takes funcions on $\mathbb{R}^n$ to functions on the dual space, $(\mathbb{R}^n)^*$ (section 1.1). One usually identifies these two spaces by choosing an inner product on $\mathbb{R}^n$, e.g., the Euclidean inner product. In the infinite–dimensional case, it is tempting to extend Bochner's theorem by letting $\mathbb{R}^n$ go to a Hilbert space and looking for a measure on this space. However, Segal [1956a, 1958] has shown that the ensuing $d\mu$ cannot be a Borel measure, since it is only finitely additive. To obtain a Borel measure, the domain of integration must be expanded to a space of distributions, its dual then being a space of test functions. Minlos' theorem (Gel'fand and Vilenkin [1964]; Glimm and Jaffe [1981]) states that if a functional $S[g]$ defined on the Schwartz space of test functions $\mathcal{S}(\mathbb{R}^s)$ satisfies appropriate conditions (positive–definiteness, normalization and continuity), a Borel probability measure $d\mu$ exists on the dual space of tempered distributions $\mathcal{S}'(\mathbb{R}^s)$ such that eq. (17) is satisfied for all $g \in \mathcal{S}(\mathbb{R}^s)$, the integration being over $\mathcal{S}'(\mathbb{R}^s)$. This resolves the problem of infinite dimansionality. In our case, however, the situation is further complicated by the fact that the space of f's over which we wish to integrate, even if it is enlarged, consists not of free functions but of solutions of the Klein–Gordon equation. This difficulty can be overcome by first applying Minlos' theorem in the momentum space representation, where $S[\hat{g}] \equiv \exp\left(-\|\hat{g}\|^2_{L^2(d\tilde{p})}\right)$ satisfies the necessary conditions for $\hat{g} \in \mathcal{S}(\mathbb{R}^s)$. This gives a probability measure $d\tilde{\mu}(\hat{f})$ on $\mathcal{S}'(\mathbb{R}^s)$. We then define the spaces

$$\begin{aligned} \mathcal{K}_0 &\equiv \left\{ g(z) = \int_{\Omega_m^+} d\tilde{p}\; e^{-izp}\, \hat{g}(\mathbf{p}) \;\mid \hat{g} \in \mathcal{S} \right\} \\ \mathcal{K}_0' &\equiv \left\{ f(z) = \int_{\Omega_m^+} d\tilde{p}\; e^{-izp}\, \hat{f}(\mathbf{p}) \;\mid \hat{f} \in \mathcal{S}' \right\}. \end{aligned} \tag{23}$$

These may be regarded as mutually dual, under the sesquilinear pairing

$$\langle\ f, g\ \rangle \equiv \langle\ \hat{f}, \hat{g}\ \rangle \equiv \int_{\Omega_m^+} d\tilde{p}\ \ \overline{\hat{f}(p)}\, \hat{g}(p), \quad \hat{f} \in \mathcal{S}',\ \hat{g} \in \mathcal{S}. \tag{24}$$

Together with $\mathcal{K}$, they form a "triplet"

$$\mathcal{K}_0 \subset \mathcal{K} \subset \mathcal{K}_0'. \tag{25}$$

We now use the map $\hat{f} \mapsto f$ to transfer the measure from $\mathcal{S}'$ to $\mathcal{K}_0'$, obtaining a probability measure $d\mu$ on $\mathcal{K}_0'$. This results, finally, in the resolution of unity

$$\int_{\mathcal{K}_0'} d\mu(f)\ |\ E^f\ \rangle\langle\ E^f\ |\ = I, \tag{26}$$

where the integral converges, as usual, in the weak operator topology. $d\mu$ is Gaussian in the sense that its restrictions to finite–dimensional cylinder sets in $\mathcal{K}_0'$ are all Gaussian measures. It is, therefore, an infinite–dimensional version of the Gaussian measure on $\mathbb{C}^s$ which gave the resolution of unity for the canonical coherent states in section 1.2.

One might well ask what is the point of insisting that the integration take place over $\mathcal{K}_0'$ rather than $\mathcal{S}'(\mathbb{R}^s)$, the momentum space representation. One reason is esthetic: The vectors E^f combine the *finite–dimensional* (particle) coherent–state representation with the *infinite–dimensional* (field) coherent–state representation. Another reason is that whereas the "sample points" $\hat{f}$ in $\mathcal{S}'(\mathbb{R}^s)$ are merely distributions, the elements f in $\mathcal{K}_0'$ are holomorphic functions, since the decaying exponential e^{-yp} dominates the singular behavior of $\hat{f}$, just as it did when $\hat{f} \in L^2(d\tilde{p})$. Note, however, that non–normalizable field coherent states E^f now enter the resolution of unity. In fact, the Hilbert space $\mathcal{K}$ has measure zero with respect to $d\mu$, since $L^2(d\tilde{p})$ has measure zero with respect to $d\tilde{\mu}$. This is remedied by the fact that only vectors h, g in the test function space $\mathcal{K}_0$ are now allowed in eq. (16).

With the resolution of unity provided by the field coherent states, the inner product in $\mathcal{H}$ can be represented by the *functional integral*

$$\begin{aligned} \langle\, \Phi \,|\, \Psi \,\rangle_{\mathcal{H}} &= \int_{\mathcal{K}_0'} d\mu(f)\, \langle\, \Phi \,|\, E^f \,\rangle \langle\, E^f \,|\, \Psi \,\rangle \\ &\equiv \int_{\mathcal{K}_0'} d\mu(f)\, \overline{\Phi[f]}\, \Psi[f]. \end{aligned} \tag{27}$$

The above construction was for a neutral scalar field. If ϕ were *charged,* its coherent states would take the form

$$E^{f,\bar{g}} = e^{\phi^*(f)+\phi(\bar{g})}\, \Psi_0, \tag{28}$$

where $\phi^*(f)$ is as before and $\phi(\bar{g})$ is an antiparticle creation operator,

$$\phi(\bar{g}) \equiv \int_{\sigma_-} d\sigma\, \phi(z)\, \bar{g}(z), \tag{29}$$

which commutes with $\phi^*(f)$. $\bar{g}(z)$ is a negative–energy solution of the Klein–Gordon equation, holomorphic in $\mathcal{T}_-$, or, equivalently, $\bar{g}(z) = \overline{g(\bar{z})}$ with $g \in \mathcal{K}_0'$. Thus we write $\bar{g} \in \overline{\mathcal{K}_0'}$. The resolution of unity for charged fields is therefore

$$\int_{\mathcal{K}_0' \times \overline{\mathcal{K}_0'}} d\mu(f,\bar{g})\, |\, E^{f,\bar{g}} \,\rangle \langle\, E^{f,\bar{g}} \,| \, = I_{\mathcal{H}}, \tag{30}$$

where $d\mu(f,\bar{g}) = d\mu(f)\, d\mu(\bar{g})$ is the tensor product of two Gaussian measures defined as above.

Finally, it is reasonable to ask whether coherent states exist for an *interpolating* field, by analogy with the interpolating particle coherent states studied in sections 5.3 and 5.6. A *necessary* condition would seem to be that the first half of the canonical commutation relations still be valid, i.e.,

$$[\phi(z), \phi(z')] = 0 \tag{31}$$

for $z, z' \in \mathcal{T}_+$, since one would like to find simultaneous eigenvectors of $\phi(z)$ for all $z \in \mathcal{T}_+$. Recall that the extended *free* neutral scalar field had the form

$$\phi(z) = \int_{\Omega_m^+} d\tilde{p}\ \left[\theta^{-izp}\, a(p) + \theta^{izp}\, a^*(p)\right] \tag{32}$$

for arbitrary $z \in \mathbb{C}^{s+1}$, and the commutation relation given by eq. (31) was due to the polarization of the positive and negative frequencies into $\mathcal{T}_+$ and $\mathcal{T}_-$, respectively. If interactions are introduced, the positive– and negative–frequency components get inextricably mixed together, hence it is highly unlikely that the above commutation relation survives. However, a *charged* free field has the form

$$\phi(z) = \int_{\Omega_m^+} d\tilde{p} \ \left[\theta^{-izp}\, a(p) + \theta^{izp}\, b^*(p)\right], \tag{33}$$

where b^* commutes with a, hence

$$[\phi(z), \phi(z')] = 0 \qquad \forall z, z' \in \mathbb{C}^{s+1}. \tag{34}$$

I believe that *this* relation does have a chance of holding for interpolating charged scalar fields. It would be a consequence, for example, of the physical requirement that the Lie algebra generated by the field has no operators which remove (or add) a double charge 2ε. This commutation relation is the *weaker half* of the free–field canonical commutation relations, the stronger half (which we do *not* assume) being that $[\phi(z), \phi(z')^*]$ is a "c–number," i.e. a multiple of the identity. If $[\phi(z), \phi(z')] = 0$ for all z and z', then it makes sense to look for common eigenvectors of all the $\phi(z)$'s, which would be coherent states of the interpolating field.

Notes

Most of the results in sections 5.2–5.5 were announced in Kaiser [1987b] and have been published in Kaiser [1987a]. An earlier attempt to describe quantized fields in complex spacetime was made in Kaiser [1980b] but was found to be unsatisfactory. The Analytic–Signal transform is further studied in Kaiser [1990c].

Segal [1963b] proposed a formulation of quantum field theory in terms of the symplectic geometry of the phase space of classical fields. This phase space corresponds, roughly, to the space $\mathcal{K}_0'$ defined in section 5.7. An attempt to study quantized fields as (operator–valued) functions on the Poincaré group–manifold has been made by Lurçat [1964]; see also Hai [1969]. As mentioned in section 4.2, this manifold may be regarded as an extended phase space which includes spin degrees of freedom in addition to position– and velocity coordinates. In

the context of *classical* field theory, this point of view has been generalized to curved spacetime by replacing the Poincaré group–manifold with the orthogonal frame bundle over a Lorentzian spacetime (Toller [1978]). These efforts have not, however, utilized holomorphy. It may be interesting to expand the point of view advocated here to a complex manifold containing the Poincaré group–manifold in order to account naturally for spin. This might result in a "total" coherent–state representation where the classical phase space coordinates range over $\mathcal{T}$ and the spin phase space coordinates range over the Riemann sphere, as in section 3.5.

I owe special thanks to R. F. Streater for many important comments and corrections in this chapter.

Chapter 6
FURTHER DEVELOPMENTS

6.1. Holomorphic Gauge theory

In this section we give a brief, not–very–technical but hopefully intuitive, discussion of gauge theory and indicate how it may be modified in order to make sense in complex spacetime. This represents work still in progress, and our account is accordingly incomplete. One obstacle is the absence, so far, of a satisfactory Lagrangian formulation. This is due in part to the fact that fields in complex spacetime are constrained since they can be derived from local fields in $\mathbb{R}^{s+1}$ (chapter 5). Our treatment is as elementary as possible, with a geometrical emphasis. We ignore global questions and work within a single chart. For more details, see Kaiser [1980a, 1981].

Gauge theory is a natural, geometric way of introducing interactions. It applies some of the ideas of General Relativity to quantum mechanics and arrives at a class of theories which are generalizations of classical electrodynamics, the latter being the simplest case. The power of Einstein's theory of gravitation owes much to the fact that it actually assumes *less*, initially, than its predecessor, Newton's theory. By dropping the assumption that spacetime is flat, we lose the ability to transport tangent vectors from one point in spacetime to another, as is done when differentiating a vector field or even finding the acceleration of a particle moving in spacetime. To regain it, we need a *connection*. We need to know how a vector transforms in going from any given point x_0 to a neighboring point along a curve $x(t)$. The tangent vector to the curve at the point x_0 is

$$X = \dot{x}^\mu \partial_\mu. \tag{1}$$

(The partial–derivative operators $\partial_\mu = \partial/\partial x^\mu$ form a basis for the tangent space at x_0; see Abraham and Marsden [1978].) An infinitesimal transport must have a linear effect, thus a vector Y at x_0 should change by

$$\delta Y = \Gamma(X) Y \, \delta t, \tag{2}$$

where $\Gamma(X)$ is a linear transformation on the tangent space at x_0. Furthermore, $\Gamma(X)$ must be linear in X since the latter is an infinitesimal (i.e., linearized) description of the curve at x_0. If $Y = Y^\nu \partial_\nu$, this gives

$$\delta Y = \dot{x}^\mu Y^\nu \Gamma(\partial_\mu) \partial_\nu \, \delta t. \tag{3}$$

Since $\Gamma(\partial_\mu)$ is a linear transformation, we have

$$\delta Y = \dot{x}^\mu Y^\nu \Gamma^\kappa_{\mu\nu} \, \partial_\kappa \, \delta t, \tag{4}$$

where $\Gamma^\kappa_{\mu\nu}$ are a set of (locally defined) funcions on spacetime, known as the *connection coefficients*. This gives the rate of change of Y due to transport along X.

Now suppose we are given a *metric* g on spacetime. (In Relativity, the "metric" is indefinite, i.e. Lorentzian rather than Riemannian; with this understood, we continue to call it a metric.) If two vectors are transported along a curve, then their inner product must not change since it is a scalar. This gives a relation betweeen Γ and g which determines the symmetric part (with respect to exchange of X and Y) of Γ. The antisymmetric part is the *torsion*, which in the standard theory is assumed to vanish. Hence the metric uniquely determines a torsionless connection known as the *Riemannian connection.*

General Relativity (see Misner, Thorne and Wheeler [1970]) relates the Riemannian connection to Newton's gravitational potential, thus giving gravity a geometric interpretation. This is reasonable, since the connection determines the acceleration of a particle moving freely (i.e., along a geodesic) in spacetime, which is related to gravity. By assuming less to begin with, one discovers what must necessarily be added to even *compute* the acceleration, and this additional structure turns out to include gravity! In Newton's theory, gravity must be added in an ad–hoc fashion. The two theories coincide in the non–relativistic limit.

Now consider the wave function of a (scalar) quantum particle, for example a solution of the free Klein–Gordon equation, in *real* spacetime $\mathbb{R}^{s+1}$. Suppose we drop the usually implicit assumption that f can be differentiated by simply taking the difference between its values at neighboring points. Instead of regarding $f(x)$ as a complex *number*, we now regard it as a one–dimensional complex *vector* attached to x, analogous to the tangent vectors in Relativity. This

means assuming less structure to begin with, since $f(x)$ now belongs to $\mathbb{C}$ as a *vector space* rather than as an *algebra*. The complex plane attached to x will be denoted by $\mathbb{C}_x$ and is called the *fiber* at x, and the set of all fibers is called a *complex line bundle* over $\mathbb{R}^{s+1}$ (Wells [1980]). To differentiate f, we must know how it is affected by transport. The situation is similar to the one above, only now $\Gamma(X)$ must be a 1×1 complex matrix, i.e. a complex number. An infinitesimal transport gives

$$\delta f = \Gamma(X) f \, \delta t = \dot{x}^\mu \Gamma_\mu f \, \delta t, \tag{5}$$

where $\Gamma_\mu(x)$ is a complex–valued function. The total rate of change along X is therefore

$$\dot{x}^\mu \partial_\mu f + \dot{x}^\mu \Gamma_\mu f \equiv D_X f, \tag{6}$$

where D_X is called the *covariant derivative along* X. Equivalently, the differential change is

$$Df = df + \Gamma f, \qquad \text{where} \qquad \Gamma \equiv \Gamma_\mu \, dx^\mu. \tag{7}$$

The 1–form Γ is called the *connection form*. Df is the sum of a "horizontal" part df (which measures change due to the dependence of f on x) and a "vertical" part (which measures change due to transport). A *gauge transformation* is represented locally by a multiplication by a variable phase factor, i.e.

$$f(x) \mapsto e^{i\chi(x)} \, f(x). \tag{8}$$

This is a linear map on each fiber $\mathbb{C}_x$, which corresponds in Relativity to the linear map on tangent spaces induced by a coordinate transformation. In fact, since there is no longer any natural way to identify distinct fibers, a gauge transformation *is* a coordinate transformation of sorts. We therefore require that Df be invariant under gauge transformations, which implies that Γ transforms as

$$\Gamma \mapsto \Gamma - i d\chi. \tag{9}$$

Suppose now that we try to complete the analogy with Relativity by deriving Γ from a Hermitian metric on the fibers such that the scalar product of two vectors remains invariant under transport. If we assume the metric to be positive–definite, it must have the form

$$(f(x), g(x)) = \bar{f}(x) \, h(x) \, g(x), \tag{10}$$

where $h(x)$ is a positive function. It will suffice to consider the inner product of $f(x)$ with itself, i.e. the quantity

$$\rho(x) \equiv (f(x),\, f(x)) = \overline{f(x)}\, h(x)\, f(x). \tag{11}$$

We require that ρ be invariant under transport. This means that (f, f) changes only due to its dependence on x, i.e.

$$(Df,\, f) + (f,\, Df) = d(f, f). \tag{12}$$

It follows that

$$\overline{\Gamma}\, h + h\Gamma = dh, \tag{13}$$

which constrains the real part of Γ but leaves the imaginary part arbitrary. Writing $\Gamma = R + iA$, where R and A are real 1–forms, we have

$$2R = d \log h. \tag{14}$$

The real part R of the connection can be transformed away by defining $\tilde{f} = h^{1/2} f$ and $\tilde{h} \equiv 1$, which gives $\tilde{\rho} = \rho$ and

$$Df = h^{-1/2}\, (d\tilde{f} + iA\tilde{f}) \equiv h^{-1/2}\, (d + \tilde{\Gamma})\tilde{f}. \tag{15}$$

Since $\rho = \overline{\tilde{f}}\, \tilde{f}$, we may as well assume from the outset that $\rho = |f|^2$ and $\Gamma = iA$ is purely imaginary.

Note: The mapping $f \mapsto \tilde{f}$ is *not* a gauge transformation in the usual sense; in the standard gauge theory the metric is assumed to be constant ($h(x) \equiv 1$), hence only phase translations are allowed. This corresponds to having already transformed away R. It turns out that in phase space, it will be natural to admit non–constant metrics.

To make the Klein–Gordon equation invariant under gauge transformations, we now replace ∂_μ by $D_\mu = \partial_\mu + iA_\mu$. The result is

$$(D^\mu D_\mu + m^2)f = (\partial^\mu + iA^\mu)(\partial_\mu + iA_\mu)f + m^2 f = 0. \tag{16}$$

This equation was known (even before gauge theory) to be a relativistically covariant description of a Klein–Gordon particle in the presence of the electromagnetic field determined by the *vector potential* $A_\mu(x)$. Hence the connection, which describes a geometric property of the the complex line bundle, acquires a physical significance with

respect to electrodynamics, just as did the connection Γ with respect to gravitation. When f is differentiated in the usual way, it is unconsciously assumed that the connection vanishes. Coupling to an electromagnetic field then has to be put in "by hand," through the substitution $\partial_\mu \to \partial_\mu + iA_\mu$, which is known as the *minimal coupling* prescription. Gauge theory gives this ad–hoc prescription a geometric interpretation. But note that in this case, the fiber metric $h(x)$ did *not* determine the connection. This is due to the complex structure: iAf cancels in the inner product because it is imaginary. The electromagnetic field generated by the potential A is given by the 2–form

$$F = dA, \tag{17}$$

which in fact measures the non–triviality of the connection form A: if $F = 0$, then A is closed and therefore (locally) exact, i.e. it is due purely to a choice of gauge. This is analogous in Relativity to choosing an accelerating coordinate system, which gives the illusion of gravity.

Note the complementary nature of the two theories: in Relativity, the skew part of the connection, which is the torsion, is assumed to vanish. In gauge theory, the inner product becomes Hermitian, and the symmetric and antisymmetric parts correspond to its real and imaginary parts, respectively. It is the *real* part of the connection which is assumed to vanish in gauge theory. Were Γ required to be real, it could in fact be transformed away as above, giving rise to no gauge field. That is, only the *trivial* part of the connection is determined by the metric. The non–trivial (imaginary) part is arbitrary. We will see that when the theory is extended to complex spacetime, the metric *does* determine a non–trivial connection.

Gauge theory usually begins with a Lagrangian invariant under phase translations $f \mapsto e^{i\phi} f$, which form the group $U(1)$. The equations satisfied by f and A_μ are derived using variational principles (see Bleeker [1981]). There is a natural generalization where $f(x)$ is an n–dimensional complex vector. In that case, the group of phase tanslations is replaced by a non–abelian group G, usually a subgroup of $U(n)$. G is called the *gauge group*, and the correponding gauge theory is said to be non–abelian or of the Yang–Mills type. Such theories have in recent years been applied with great success to the two remaining known interactions (aside from gravity and electromagnetism), namely the *weak* and the *strong* forces, which involve nuclear matter (Appelquist et al. [1987]).

Let us now see how non–abelian gauge theory may be extended

to *complex* spacetime. (This will include the abelian case of electrodynamics when $n = 1$.) Consider a field f on complex spacetime, say on the double tube $\mathcal{T}$, whose values are n–dimensional complex vectors. The set of all possible values at $z \in \mathcal{T}$ is a complex vector space $F_z \approx \mathbb{C}^n$ called the *fiber* at z. The collection of all fibers is called a *vector bundle*. We assume that this bundle is holomorphic (Wells [1980]), so that holomorphic sections $z \mapsto f(z) \in F_z$, represented locally by holomorphic vector–valued functions, make sense. Upon transport along a curve $z(t)$ having the complex tangent vector Z, f changes by

$$\delta f = \theta(Z) f \, \delta t, \tag{18}$$

where $\theta(Z)$ is a linear map on each fiber. The total differential change is

$$Df = (d + \theta) \, f. \tag{19}$$

If $z = x - iy$, then

$$d = \partial + \bar{\partial}, \tag{20}$$

where

$$\begin{aligned} \partial &= dz^\mu \, \partial_\mu = \frac{1}{2} \, dz^\mu \left(\frac{\partial}{\partial x^\mu} + i \frac{\partial}{\partial y^\mu} \right) \\ \bar{\partial} &= d\bar{z}^\mu \, \bar{\partial}_\mu = \frac{1}{2} \, d\bar{z}^\mu \left(\frac{\partial}{\partial x^\mu} - i \frac{\partial}{\partial y^\mu} \right). \end{aligned} \tag{21}$$

Since a general tangent vector has the form

$$Z = Z_\mu \, \partial_\mu + Z_{\bar{\mu}} \, \bar{\partial}_\mu, \tag{22}$$

we have

$$\theta(Z) = \theta_\mu \, dz^\mu + \theta_{\bar{\mu}} \, d\bar{z}^\mu, \tag{23}$$

where $\theta_\mu = \theta(\partial_\mu)$ and $\theta_{\bar{\mu}} = \theta(\bar{\partial}_\mu)$.

Let us now try again to derive the connection from a fiber metric, as we have failed to in the case of real spacetime. A positive–definite metric on the fibers F_z must have the form

$$(f(z), \, g(z)) = f(z)^* \, h(z) \, g(z) \tag{24}$$

where $h(z)$ is a positive–definite matrix. Again, it will suffice to consider the (squared) fiber norm

$$\rho(z) = (f(z),\, f(z)). \tag{25}$$

We define a *holomorphic gauge transformation* to be a map of the form

$$\begin{aligned} f(z) &\mapsto f'(z) = \chi(z)^{-1}\, f(z), \\ h(z) &\mapsto h' = \chi(z)^*\, h(z)\, \chi(z), \end{aligned} \tag{26}$$

where $\chi(z)$ is an invertible $n \times n$ matrix–valued holomorphic function. Clearly ρ is invariant under holomorphic gauge transformations. The corresponding gauge group acting on a single fiber F_z is the general linear group $G = GL(n, \mathbb{C})$, which includes the usual gauge group $U(n)$. However, analyticity correlates the values of $\chi(z)$ at different fibers. Invariance of the inner product under transport gives

$$(Df, f) + (f, Df) = d(f, f), \tag{27}$$

from which we obtain the matrix equation

$$\theta^* h + h\theta = dh = \bar\partial h + \partial h. \tag{28}$$

As in the case of real spacetime, this only determines the Hermitian part (relative to the metric) of θ. But if we make the *Ansatz*

$$h\theta = \partial h, \tag{29}$$

then the resulting connection $\theta = h^{-1}\partial h$ satisfies the above constraint.

We now show that in general, the connection θ is non–trivial, i.e. cannot be transformed away by a holomorphic gauge transformation. Under such a transformation, θ becomes

$$\begin{aligned} \theta' &= (\chi^* h \chi)^{-1} \partial (\chi^* h \chi) \\ &= \chi^{-1} h^{-1} \partial h \chi + \chi^{-1} \partial \chi \\ &= \chi^{-1} \theta \chi + \chi^{-1} \partial \chi, \end{aligned} \tag{30}$$

since $\partial \chi^* = 0$ by analyticity. It follows that

$$D' f' \equiv (d + \theta') f' = \chi^{-1} D(\chi f'), \tag{31}$$

hence

$$(D')^2 f' = \chi^{-1} D^2 (\chi f'). \tag{32}$$

If θ were trivial, then for some gauge we would have $\theta' = 0$, hence $(D')^2 = d^2 = 0$, so $D^2 = 0$. But

$$\begin{aligned} D^2 f &= (d+\theta)(d+\theta)\, f = d(\theta\, f) + \theta \wedge df + \theta \wedge \theta f \\ &= (d\theta)\, f + \theta \wedge \theta\, f \equiv \Theta\, f, \end{aligned} \tag{33}$$

where the 2–form

$$\Theta = d\theta + \theta \wedge \theta = \bar\partial\theta + \partial\theta + \theta \wedge \theta \tag{34}$$

is the *curvature form* of the connection θ, analogous to the Riemann curvature tensor in Relativity. Using the matrix equation

$$dh^{-1} = -h^{-1} dh \cdot h^{-1} \tag{35}$$

and $\partial^2 = \bar\partial^2 = \partial\bar\partial + \bar\partial\partial = 0$, we find

$$\partial\theta + \theta \wedge \theta = \left(-h^{-1}\partial h \cdot h^{-1}\right) \wedge \partial h + \left(h^{-1}\partial h\right) \wedge \left(h^{-1}\partial h\right) = 0. \tag{36}$$

This is an *integrability condition* for θ, being a consequence of the fact that θ can be derived from h. One could say that h is a "potential" for θ. Therefore the quadratic term cancels in eq. (34) and the curvature form reduces to

$$\Theta = \bar\partial\theta. \tag{37}$$

Hence if h is such that $\bar\partial\left(h^{-1}\partial h\right) \neq 0$, then the connection is non–trivial. The form Θ is the complex spacetime version of a *Yang–Mills field*, and θ corresponds to the Yang–Mills potential.

For $n = 1$, θ and Θ are the complex spacetime versions of the electromagnetic potential and the electromagnetic field, respectively. Since $h(z)$ is a positive function, it may be written as

$$h(z) = e^{-\phi(z)} \tag{38}$$

where $\phi(z)$ is real. Then

$$\theta = -\partial\phi, \qquad \Theta = -\bar\partial\partial\theta. \tag{39}$$

To relate θ and Θ to the electromagnetic potential A and the electromagnetic field F, one performs a non–holomorphic gauge transformation similar to that in eq.(15): let

$$\tilde{f}(z) = e^{-\phi(z)/2} f(z), \qquad \tilde{h}(z) \equiv 1. \tag{40}$$

Then the transformed potential becomes purely imaginary,

$$\tilde{\theta} = -\frac{i}{2} \frac{\partial \phi}{\partial y^\mu} dx^\mu, \tag{41}$$

giving

$$A_\mu(z) = -\frac{1}{2} \frac{\partial \phi}{\partial y^\mu} \tag{42}$$

for the complex spacetime version of the electromagnetic potential. Note that although $A_\mu(z)$ is a pure gradient in the y–direction, the corresponding electromagnetic field is *not* trivial, since

$$\begin{aligned} F_{\mu\nu}(z) &\equiv \frac{\partial A_\mu}{\partial x^\nu} - \frac{\partial A_\nu}{\partial x^\mu} \\ &= \frac{1}{2} \left[\frac{\partial^2 \phi}{\partial x^\mu \partial y^\nu} - \frac{\partial^2 \phi}{\partial x^\nu \partial y^\mu} \right] \end{aligned} \tag{43}$$

need not vanish, in general.

Incidentally, there is an intriguing similarity between the inner product using the fiber metric,

$$\langle\, f \,|\, g \,\rangle = \int_\sigma d\mu(z)\, \overline{f(z)}\, e^{-\phi(z)}\, g(z), \tag{44}$$

and that in Onofri's holomorphic coherent states representation (sec. 3.4, eq. (62)). Note that our Θ coincides with Onofri's symplectic form $-\omega$. This possible connection remains to be explored.

Remarks.

1. The relation between the Yang–mills potential A and the Yang–Mills field F in *real* spacetime is $F = dA + A \wedge A$, which is *quadratic* in the non–abelian case ($n > 1$), since the wedge product $A \wedge A$ involves matrix multiplication. In our case, however, the connection satisfies the integrability condition given by eq. (36), hence the quadratic term cancels and the relation becomes *linear*, just as it is normally in the abelian case. The non–holomorphic gauge transformation $f \mapsto \tilde{f}$ in eq. (40) can be generalized to the non–abelian case as follows: Since $h(z)$ is a postive matrix, it can be written as $h(z) = k(z)^* k(z)$, where $k(z)$ is an

$n \times n$ matrix. ($k(z)$ need not be Hermitian; the holomorphic case $\bar{\partial} k = 0$ corresponds to a "pure gauge" field, i.e. $\Theta = 0$.) Setting $\tilde{f}(z) = k(z) f(z)$ and $\tilde{h}(z) \equiv 1$ brings us to the *unitary gauge*, where the new gauge transformations are given by unitary matices $\tilde{f}(z) \mapsto U(z)\tilde{f}(z)$. This amounts to a *reduction* of the gauge group from $GL(n, \mathbb{C})$ to $U(n)$. In the unitary gauge, the relation between the connection and the curvature becomes non–linear, as it is in the usual Yang–Mills theory. See Kaiser [1981] for details.

2. Θ has a symmetric real part and an antisymmetric imaginary part. The antisymmetric part corresponds to the usual Yang–Mills field, whereas the symmetric part does not seem to have an obvious counterpart in real spacetime.
3. In complex differential geometry, $\theta = h^{-1}\partial h$ is known as the *canonical connection of type* (1, 0) determined by h. (Wells [1980]). The functions f are assumed to be (local representations of) holomorphic sections of the vector bundle. We do not make this assumption, since it appears that analyticity may be lost in the presence of interactions. However, it is possible that $f(z)$ is holomorphic, and it is the non–holomorphic gauge transformation $f \mapsto \tilde{f}$ which spoils the analyticity. That is, the non–analytic part of the theory may be all contained in the fiber metric h.

6.2 Windowed X–Ray Transforms: Wavelets Revisited

In this final section we generalize the idea behind the Analytic–Signal transform (section 5.2) and arrive at an n–dimensional version of the Wavelet transform (chapters 1 and 2). In a certain sense, relativistic wave functions and fields in complex spacetime may be regarded as *generalized wavelet transforms* of their counterparts in real spacetime. This is related to the fact, mentioned earlier, that relativistic windows shrink in the direction of motion, due to the Lorentz contraction associated with the hyperbolic geometry of spacetime. This contraction is like the compression associated with wavelets.

Other generalizations of the wavelet transform to more than one dimension have been studied (see Malat [1987]), but they are usually obtained from the one–dimensional one by taking tensor products, hence are not natural with respect to symmetries of $\mathbb{R}^n$ (such as rotations or Lorentz transformations), and consequently do not lends themselves to analysis by group–theoretic methods. The generaliza-

tion proposed here, which we call a *windowed X–Ray transform,* does not assume any prefered directions, hence it respects all symmetries of $\mathbb{R}^n$. For example, the transforms of functions over real spacetime $\mathbb{R}^{s+1}$ will transform naturally under the Poincaré group. If these functions form a representation space for $\mathcal{P}_0$ (whether irreducible, as in the case of a free particle, or reducible, as in the case of systems of interacting fields), then so do their transforms.

Let us start directly with the windowed X–Ray transform (Kaiser [1990b]). Fix a window function $h\colon \mathbb{R} \to \mathbb{C}$, which will play a role similar to a "basic wavelet." For a given (sufficiently well–behaved) function f on $\mathbb{R}^n$, define f_h on $\mathbb{R}^n \times \mathbb{R}^n$ by

$$f_h(x,y) = \int_{-\infty}^{\infty} dt\, \overline{h(t)}\, f(x+ty). \tag{1}$$

Note that for $n = 1$ and $y \neq 0$,

$$\begin{aligned} f_h(x,y) &= |y|^{-1} \int dt'\, \overline{h\left(\frac{t'-x}{y}\right)}\, f(t') \\ &= |y|^{-1/2}\, (Wf)(x,y), \end{aligned} \tag{2}$$

where Wf is the usual Wavelet transform based on the affine group $A = \mathbb{R} \circledS \mathbb{R}^*$, defined in section 1.6. On the other hand, for arbitrary n but

$$h(t) = \frac{1}{2\pi(1-it)}, \tag{3}$$

f_h coincides with the Analytic–Signal transform defined in section 5.2. The name "windowed X–Ray transform" derives from the fact that for the choice $h \equiv 1$, $f_h(x,y)$ is simply the integral of f along the line $x(t) = x + ty$, and if $|y| = 1$ this is known as the *X–Ray transform* (Helgason [1984]), due to its applications in tomography (Herman [1979]).

Returning to n dimensions and an arbitrary window function $h(t)$, we wish to know, first of all, whether and how f can be reconstructed from f_h. Note that the transform is trivial for $y = 0$, since

$$f_h(x,0) = f(x) \int dt\, \overline{h(t)}, \tag{4}$$

so we rule out $y = 0$ and let y range over $\mathbb{R}^n_* \equiv \mathbb{R}^n \backslash \{0\}$. Note also that f_h has the following *dilation property* for $a \neq 0$:

$$f_h(x,ay) = \int dt\,|a|^{-1}\overline{h(t/a)}\,f(x+ty) = f_{h_a}(x,y), \tag{5}$$

where $h_a(t) \equiv |a|^{-1}\,h(t/a)$.

To reconstruct f, begin by formally substituting the Fourier representation of f into f_h:

$$\begin{aligned} f_h(x,y) &= \int dt \int d^n p\, e^{-2\pi i p(x+ty)}\,\overline{h(t)}\,\hat{f}(p) \\ &= \int d^n p\, e^{-2\pi i p x}\,\overline{\hat{h}(py)}\,\hat{f}(p) \\ &\equiv \langle\, \hat{h}_{x,y} \,|\, \hat{f}\,\rangle_{L^2} = \langle\, h_{x,y} \,|\, f\,\rangle_{L^2}, \end{aligned} \tag{6}$$

where $\hat{h}_{x,y}$ is defined by

$$\hat{h}_{x,y}(p) = e^{2\pi i p x}\,\hat{h}(py), \tag{7}$$

so that

$$h_{x,y}(x') = \int d^n p\, e^{-2\pi i p(x'-x)}\,\hat{h}(py). \tag{8}$$

Note that $\hat{h}_{x,y}$, and hence also $h_{x,y}$, is not square–integrable for $n > 1$, since its modulus is constant along directions orthogonal to y.* This means that eq. (6) will not make sense for arbitrary $f \in L^2(\mathbb{R}^n)$. We therefore assume, initially, that f is a test function, say it belongs to the Schwartz space $\mathcal{S}(\mathbb{R}^n)$. Then eq. (6) makes sense with $\langle\ |\ \rangle$ as the (sesquilinear) pairing between distributions and test functions. If h is also sufficiently well–behaved, then we can substitute

$$\hat{h}(py) = \int d\nu\,\hat{h}(\nu)\,\delta(py-\nu) \tag{9}$$

and change the order of integration in eq. (8), obtaining

$$h_{x,y}(x') = \int d\nu\,\hat{h}(\nu) \int d^n p\, e^{-2\pi i p(x'-x)}\,\delta(py-\nu). \tag{10}$$

* So far, we have *not* assumed any metric structure on $\mathbb{R}^{s+1}$. Recall (sec. 1.1) that the natural domain of $\hat{f}(p)$ is the *dual* space $(\mathbb{R}^n)^*$. "Orthogonal" here means that $p(y) = 0$ as a linear functional. This will be important when $\mathbb{R}^n$ is spacetime with its Minkowskian metric.

To reconstruct f, we look for a resolution of unity in terms of the vectors $h_{x,y}$ (chapter 1). That is, we need a measure $d\mu(x,y)$ on $\mathbb{R}^n \times \mathbb{R}^n_*$ such that

$$\int d\mu(x,y)\,|f_h(x,y)|^2 = \|f\|^2_{L^2}. \tag{11}$$

For then the map $T\colon f \mapsto f_h$ is an isometry onto its range in $L^2(d\mu)$, and polarization gives

$$\langle\, g \,|\, f \,\rangle = \langle\, Tg \,|\, Tf \,\rangle = \langle\, g \,|\, T^*Tf \,\rangle, \tag{12}$$

showing that $f = T^*(Tf)$ in $L^2(\mathbb{R}^n)$, which is the desired reconstruction formula. There are various ways to obtain a resolution of unity, since f is actually overdetermined by f_h, i.e. giving the values of f_h on all of $\mathbb{R}^n \times \mathbb{R}^n_*$ amounts to "oversampling," so f_h will have to satisfy a consistency condition. We have seen several examples of this in the study of the windowed Fourier transform (section 1.5) and the one–dimensional wavelet transform (section 1.6 and chapter 2), where *discrete* subframes were obtained starting with a continuous resolution of unity. However, for $n > 1$, there are other options than discrete subframes, as we will see. In the spirit of the one–dimensional wavelet transform, our first resolution of unity will involve an integration over all of $\mathbb{R}^n \times \mathbb{R}^n_*$. Note that

$$f_h(x,y) = \left(\overline{\hat{h}(py)}\,\hat{f}\right)\check{}\,(x), \tag{13}$$

so Plancherel's theorem gives

$$\int d^n x\,|f_h(x,y)|^2 = \int d^n p\,|\hat{h}(py)|^2\,|\hat{f}(p)|^2. \tag{14}$$

We therefore need a measure $d\rho(y)$ on $\mathbb{R}^n_*$ such that

$$H(p) \equiv \int d\rho(y)\,|\hat{h}(py)|^2 \equiv 1 \quad \text{for almost all } p. \tag{15}$$

The solution is simple: Every $p \neq 0$ can be transformed to $q \equiv (1,0,\ldots 0)$ by a *dilation and rotation* of $\mathbb{R}^n$. That is, the orbit of q (in Fourier space) under dilations and rotations is all of $\mathbb{R}^n_*$. Thus we choose $d\rho$ to be invariant under rotations and dilations, which gives

$$d\rho(y) = N|y|^{-n}d^n y, \tag{16}$$

where N is a normalization constant and $|y|$ is the Euclidean norm of y. Then for $p \neq 0$,

$$\begin{aligned} H(p) = H(q) &= N \int |y|^{-n} d^n y \, |\hat{h}(y_1)|^2 \\ &= N \int dy_1 \, |\hat{h}(y_1)|^2 \int \frac{dy_2 \cdots dy_n}{(y_1^2 + \cdots + y_n^2)^{n/2}}. \end{aligned} \tag{17}$$

But a straightforward computation gives

$$\int \frac{dy_2 \cdots dy_n}{(y_1^2 + \cdots + y_n^2)^{n/2}} = \frac{\pi^{n/2}}{|y_1| \, \Gamma(n/2)}. \tag{18}$$

This shows that the measure $d\mu(x, y) \equiv d^n x \, d\rho(y)$ gives a resolution of unity if and only if

$$c_h \equiv \int \frac{d\xi}{|\xi|} \, |\hat{h}(\xi)|^2 < \infty, \tag{19}$$

which is precisely the *admissibility condition* for the usual (one–dimensional) Wavelet transform (section 1.6). If h is admissible, the normalization constant is given by

$$N = \frac{\Gamma(n/2)}{\pi^{n/2} \, c_h} \tag{20}$$

and the reconstruction formula is then

$$f(x') = (T^* T f)(x') = N \int \frac{d^n x \, d^n y}{|y|^n} \, \overline{h_{x,y}(x')} \, f_h(x, y). \tag{21}$$

The *sense* in which this formula holds depends, of course on the behavior of f. The class of possible f's, in turn, depends on the choice of h. A rigorous analysis of these questions is not easy, and will not be attempted here. Note that in spite of the factor $|y|^n$ in the denominator, there is no problem at $y = 0$ since

$$f_h(x, 0) = \overline{\hat{h}(0)} \, f(x) = 0 \tag{22}$$

by the admissibility condition. The behavior of f_h for small y can be analized using the dilation property, since eq. (5) implies that for $\lambda > 0$,

$$f_h(x, y/\lambda) = \int dt\, \lambda \overline{h(\lambda t)}\, f(x + ty). \tag{23}$$

Thus if $h(t)$ decays rapidly, say if

$$\lambda h(\lambda t) \to 0 \quad \text{as} \quad \lambda \to \infty, \tag{24}$$

then we expect $f_h(x, y/\lambda) \to 0$ as $\lambda \to \infty$.

Since eq. (11) holds for admissible h, we can now allow $f \in L^2(\mathbb{R}^n)$. We would like to characterize the *range* $\Re_T$ of the map T: $f \mapsto f_h$ from $L^2(\mathbb{R}^n)$ to $L^2(d\mu)$. The relation

$$f_h(x, y) = \langle\, h_{x,y} \,|\, f \,\rangle_{L^2} \tag{25}$$

shows that $h_{x,y}$ acts like an *evaluation map* taking $f_h \in L^2(d\mu)$ to its "value" at (x, y). These linear maps on $\Re_T$ are, however, not bounded if $n > 1$, since then $h_{x,y}$ is not square–integrable. (In general, the "value" of f_h at a point may be undefined.) Hence $\Re_T$ is *not* a reproducing–kernel Hilbert space (chapter 1). But in any case, the distributional kernel

$$\begin{aligned} K(x, y;\, x', y') &\equiv \langle\, h_{x,y} \,|\, h_{x',y'} \,\rangle \\ &= \int d^n p\, e^{-2\pi i p(x - x')}\, \overline{\hat{h}(py)} \hat{h}(py') \end{aligned} \tag{26}$$

represents the orthogonal projection from $L^2(d\mu)$ onto $\Re_T$. Thus a given function in $L^2(d\mu)$ belongs to $\Re_T$ if and only if it satisfies the *consistency condition*

$$g(x, y) = \int d\mu(x', y')\, K(x, y;\, x', y')\, g(x', y'), \tag{27}$$

where the integral is the symbolic representation of the action of K as a distribution.

Remarks.

1. For $n = 1$, the reconstruction formula is identical with the one for the continuous one–dimensional wavelet transform Wf, since by Eq. (2),

$$\int \frac{dx\, dy}{|y|}\, |f_h(x, y)|^2 = \int \frac{dx\, dy}{y^2}\, |(Wf)(x, y)|^2 \,. \tag{28}$$

2. In deriving the resolution of unity and the related reconstruction formula, we have tacitly identified $\mathbb{R}^n$ as a Euclidean space, i.e. we have equipped it with the Euclidean metric and identified the pairing px in the Fourier transform as the inner product. The exact place where this assumption entered was in using the rotation group plus dilations to obtain $\mathbb{R}^n_*$ from the single vector q, since rotations presume a metric.

Having established f_h as a generalization of the one–dimensional wavelet transform, let us now investigate it in its own right. First, note that for $n = 1$ there were only two simple types of candidates for generalized frames of wavelets: (a) all continuous translations and dilations of the basic wavelet, or (b) a discrete subset thereof. For $n > 1$, any choice of a discrete subset of vectors $h_{x,y}$ spoils the invariance under continuous symmetries such as rotations, and it is therefore not obvious how to use the above group–theoretic method to find discrete subframes. In fact, the discrete subsets $\{(a^m, na^m b)\}$ which gave frames of wavelets in section 1.6 and chapter 2 do *not* form subgroups of the affine group. One of the advantages of using tensor products of one–dimensional wavelets is that they do generate discrete frames for $n > 1$, though sacrificing symmetry. However, other options exist for choosing generalized (continuous) subframes when $n > 1$, and one may adapt one's choice to the problem at hand. Such choices fall between the two extremes of using *all* the vectors $h_{x,y}$ and merely summing over a discrete subset, as seen in the examples below.

1. The X–Ray Transform

The usual X–Ray transform is obtained by choosing $h(t) \equiv 1$, which is not admissible in the above sense; hence the above "wavelet" reconstruction fails. The reason is easy to see: Note that now f_h has the following symmetries:

$$\begin{aligned} f_h(x, ay) &= |a|^{-1} f_h(x, y) \quad \forall\, a \in \mathbb{R}^* \\ f_h(x + sy, y) &= f_h(x, y) \quad \forall s \in \mathbb{R}. \end{aligned} \tag{29}$$

Together, these equations state that f_h depends only on the line of integration and not on the way it is parametrized. The first equation shows that integration over all $y \neq 0$ is unnecessary as well as undesirable, and it suffices to integrate over the unit sphere $|y| = 1$. The second equation shows that for a given y, it is (again) unnecessary and undesirable to integrate over all x, and it suffices to integrate over the

hyperplane orthogonal to y. The set of all such (x, y) does, in fact, correspond to the set of all lines in $\mathbb{R}^n$, and the corresponding set of $h_{x,y}$'s forms a continuous frame which gives the usual reconstruction formula for the X–Ray transform (Helgason [1984]). The moral of the story is that sometimes, inadmissibility in the "wavelet" sense carries a message: Reduce the size of the frame.

2. The Radon Transform

Next, choose $\nu \in \mathbb{R}$ and

$$h_\nu(t) = e^{-2\pi i\nu t}. \tag{30}$$

Like the previous function, this one is inadmissible, hence the "wavelet" reconstruction fails. Again, this can be corrected by understanding the reason for inadmissibility. Eq. (6) now gives

$$f_{h_\nu}(x, y) = \int d^n p\, e^{-2\pi ipx}\, \delta(py - \nu)\, \hat{f}(p). \tag{31}$$

For any $a \neq 0$, we have

$$f_{h_\nu}(x, ay) = |a|^{-1} f_{h_\omega}(x, y), \tag{32}$$

where $\omega = \nu/a$. Hence it suffices to restrict the y–integration to the unit sphere, provided we also integrate over $\nu \in \mathbb{R}$. Also, for any $\tau \in \mathbb{R}$,

$$\begin{aligned} f_{h_\nu}(x + \tau y, y) &= \int d^n p\, e^{-2\pi ipx}\, e^{-2\pi i\tau py}\, \delta(py - \nu)\, \hat{f}(p) \\ &= e^{-2\pi i\tau\nu}\, f_{h_\nu}(x, y). \end{aligned} \tag{33}$$

Fixing $x = 0$, the function

$$(R\hat{f})(y, \nu) = f_{h_\nu}(0, y) \tag{34}$$

is called the *Radon transform* of $\hat{f}$ (Helgason [1984]). It may be regarded as being defined on the set of all hyperplanes in the Fourier space $(\mathbb{R}^n)^*$, and $\hat{f}$ can be reconstructed by integrating over the set of these hyperplanes.

3. The Fourier–Laplace Transform

Now consider

$$h(t) = \frac{1}{2\pi i(t-i)} \tag{35}$$

which gives rise to the Analytic–Signal transform. (We have adopted a slightly different sign convention than is sec. 5.2. Also, note that we have re–inserted a factor of 2π in the exponent in the Fourier transform, which simplifies the notation.) Then $\hat{h}$ is the exponential step function (sec. 5.2)

$$\hat{h}(\xi) = \theta(\xi)\, e^{-2\pi\xi} = \theta^{-2\pi\xi}, \tag{36}$$

and eq. (6) reads

$$\begin{aligned} f_h(x,y) &= \int d^n p\, \theta(py)\, e^{-2\pi ip(x-iy)}\, \hat{f}(p) \\ &= \int_{M_y} d^n p\, e^{2\pi ip(x-iy)}\, \hat{f}(p), \end{aligned} \tag{37}$$

where M_y is the half–space $\{p \,|\, py > 0\}$. This is the *Fourier–Laplace transform* of $\hat{f}$ in M_y. For $n = 1$ and $y > 0$, it reduces to the usual Fourier–Laplace transform.

This h, too, is not admissible. $f(x)$ can be recovered simply by letting $y \to 0$, and $f_h(x,y)$ may be regarded as a *regularization* of $f(x)$. If the support of $\hat{f}$ is contained in some closed convex cone $\Gamma^* \subset (\mathbb{R}^n)^*$, then $f_h(x,y) \equiv f(x-iy)$ is holomorphic in the tube $\mathcal{T}_\Gamma$ over the cone Γ *dual* to Γ^*, i.e.

$$\begin{aligned} \Gamma &= \{y \in \mathbb{R}^n \,|\, py > 0 \quad \forall\, p \in \Gamma^*\} \\ \mathcal{T}_\Gamma &= \{x - iy \in \mathbb{C}^n \,|\, y \in \Gamma\}. \end{aligned} \tag{38}$$

(Note that no metric has been assumed.) In that case, $f(x)$ is a *boundary value* of $f(x-iy)$. This forms the background for the theory of Hardy spaces (Stein and Weiss [1971]). We have encountered a similar situation when $\mathbb{R}^n$ was spacetime ($n = s+1$), $\Gamma^* = \overline{V}_+$, and $f(x)$ was a positive–energy solution of the Klein–Gordon equation; then $\Gamma = V'$ and $\mathcal{T}_\Gamma = \mathcal{T}_+$. But in that case, $f(x)$ was not in $L^2(\mathbb{R}^{s+1})$ due to the conservation of probability. There it was unnecessary and undesirable to integrate $|f(z)|^2$ over all of $\mathcal{T}_+$ since it was determined by its values on any phase space $\sigma_+ \subset \mathcal{T}_+$, and reconstruction was then achieved by integrating over σ_+ (chapter 4).

As seen from these examples, the windowed X–Ray transform has the remarkable feature of being related to most of the "classical" integral transforms: The X–Ray, Radon and Fourier–Laplace transforms. Since the Analytic–Signal transform is a close relative of the multivariate Hilbert transform H_y (sec. 5.2), we may also add H_y to this collection.

REFERENCES

Abraham, R. and Marsden, J. E.
[1978] *Foundations of Mechanics,* Benjamin/Cummings, Reading, MA.
Abramowitz, M. and Stegun, L. A., eds.
[1964] *Handbook of Mathematical Functions,* National Bureau of Standards, Applied Math. Series **55**.
Ali, T. and Antoine, J.–P.
[1989] Ann. Inst. H. Poincaré **51**, 23.
Applequist, T. Chodos, A. and Freund, P. G. O., eds.
[1987] *Modern Kaluza–Klein Theories,* Addison–Wesley, Reading, MA.
Aronszajn, N.
[1950] Trans. Amer. Math. Soc. **68**, 337.
Aslaksen, E. W. and Klauder, J. R.
[1968] J. Math. Phys. **9**, 206.
[1969] J. Math. Phys. **10**, 2267.
Bargmann, V.
[1954] Ann. Math. **59**, 1.
[1961] Commun. Pure Appl. Math. **14**, 187.
Bargmann, V., Butera, P., Girardello, L. and Klauder, J. R.
[1971] Reps. Math. Phys. **2**, 221.
Barut, A.O. and Malin, S.
[1968] Revs. Mod. Phys. **40**, 632.
Battle, G.
[1987] Commun. Math. Phys. **110**, 601.
Berezin, F. A.
[1966] *The Method of Second Quantization,* Academic Press, New York.
Bergman, S.
[1950] *The Kernel Function and Conformal Mapping,* Amer. Math. Soc., Providence.
Bialinicki–Birula, I. and Mycielski, J.
[1975] Commun. Math. Phys. **44**, 129.
Bleeker, D.
[1981] *Gauge Theory and Variational Principles,* Addison– Wesley, Reading, MA.
Bohr, N. and Rosenfeld, L.
[1950] Phys. Rev. **78**, 794.
Bott, R.
[1957] Ann. Math. **66**, 203.

Carey, A. L.
[1976] Bull. Austr. Math. Soc. **15**, 1.
Daubechies, I.
[1988a] "The wavelet transform, time–frequency localization and signal analysis," to be published in IEEE Trans. Inf. Thy.
[1988b] Commun. Pure Appl. Math. XLI, 909.
De Bièvre, S.
[1989] J. Math. Phys. **30**, 1401.
Dell'Antonio, G. F.
[1961] J. Math. Phys. **2**, 759.
Dirac, P. A. M.
[1943] *Quantum Electrodynamics,* Commun. of the Dublin Inst. for Advanced Studies series A, #1.
[1946] *Developments in Quantum Electrodynamics,* Commun. of the Dublin Inst. for Advanced Studies series A, #3.
Dufflo, M. and Moore, C. C.
[1976] J. Funct. Anal. **21**, 209.
Dyson, F. J.
[1956] Phys. Rev. **102**, 1217.
Fock, V. A.
[1928] Z. Phys. **49**, 339.
Gel'fand, I. M., Graev, M. I. and Vilenkin, N. Ya.
[1966] Generalized Functions, vol 5, Academic Press, New York.
Gel'fand, I. M. and Vilenkin, N. Ya.
[1964] Generalized Functions, vol 4, Academic Press, New York.
Gerlach, B., Gomes, D. and Petzold, J.
[1967] Z. Physik **202**, 401.
Glauber, R. J.
[1963a] Phys. Rev. **130**, 2529.
[1963b] Phys. Rev. **131**, 2766.
Glimm, J. and Jaffe, A.
[1981] *Quantum Physics,* Springer, New York.
Green, M. W., Schwarz, J. H. and Witten, E.
[1987] *Superstring Theory,* vols. I, II, Cambridge Univ. Press, Cambridge.
Greenberg, O. W.
[1962] J. Math. Phys. **3**, 859.
Guillemin, V. and Sternberg, S.
[1984] *Symplectic Techniques in Physics,* Cambridge Univ. Press, Cambridge.
Hai, N. X.
[1969] Commun. Math. Phys. **12**, 331.

Halmos, P.
[1967] *A Hilbert Space Problem Book,* Van Nostrand, London.
Hegerfeldt, G. C.
[1985] Phys. Rev. Lett. **54**, 2395.
Helgason, S.
[1962] *Differential Geometry and Symmetric Spaces,* Academic Press, NY.
[1978] *Differential Geometry, Lie Groups, and Symmetric Spaces,* Academic Press, New York.
[1984] *Groups and Geometric Analysis,* Academic Press, New York.
Henley, E. M. and Thirring, W.
[1962] *Elementary Quantum Field Theory,* McGraw–Hill, New York.
Herman, G. T., ed.
[1979] *Image Reconstruction from Projections: Implementation and Applications,* Topics in Applied Physics **32**, Springer New York.
Hermann, R.
[1966] *Lie Groups for Physicists,* Benjamin, New York.
Hille, E.
[1972] Rocky Mountain J. Math. **2**, 319.
Inönü, E. and Wigner, E. P.
[1953] Proc. Nat. Acad. Sci., USA, **39**, 510.
Itzykson, C. and Zuber, J.–B.
[1980] *Quantum Field Theory,* McGraw–Hill, New York.
Jost, R.
[1965] *The General Theory of Quantized Fields,* Amer. Math. Soc., Providence.
Kaiser, G.
[1974] "Coherent States and Asymptotic Representations," in Proc. of Third Internat. Colloq. on Group–Theor. Meth. in Physics, Marseille.
[1977a] " Relativistic Coherent–State Representations," in *Group–Theoretical Methods in Physics,* R. Sharp and B. Coleman, eds., Academic Press, New York.
[1977b] J. Math. Phys. **18**, 952.
[1977c] *Phase–Space Approach to Relativistic Quantum Mechanics,* Ph. D. thesis, Mathematics Dept, Univ. of Toronto.
[1977d] "Phase–Space Approach to Relativistic Quantum Mechanics," in proc. of Eighth Internat. Conf. on General Relativity, Waterloo, Ont.
[1978a] J. Math. Phys. **19**, 502.

[1978b] "Hardy spaces associated with the Poincaré group," in *Lie Theories and Their Applications,* W. Rossmann, ed., Queens Univ. Press, Ont.
[1978c] "Local Fourier Analysis and Synthesis," Univ. of Lowell preprint (unpublished).
[1979] Lett. Math. Phys. **3**, 61.
[1980a] "Holomorphic gauge theory," in *Geometric Methods in Mathematical Physics,* G. Kaiser and J. E. Marsden, eds., Lecture Notes in Math. **775**, Springer, Heidelberg.
[1980b] "Relativistic quantum theory in complex spacetime," in *Differential Geometrical Methods in Mathematical Physics,* P.L. Garcia, A. Perez–Rendon and J.–M. Souriau, eds., Lecture Notes in Math. **836**, Springer, Berlin.
[1981] J. Math. Phys. **22**, 705.
[1984a] "A sampling theorem for signals in the joint time– frequency domain," Univ. of Lowell preprint (unpublished).
[1984b] "On the relation between the formalisms of Prugovečki and Kaiser," Univ. of Lowell preprint, presented as a poster at the VII-th International Congress of Mathematical Physics, in *Mathematical Physics VII,* W. E. Brittin, K. E. Gustafson and W. Wyss, eds., North–Holland, Amsterdam
[1987a] Ann. Phys. **173**, 338.
[1987b] "A phase–space formalism for quantized fields," in *Group–Theoretical Methods in Physics,* R. Gilmore, ed., World Scientific, Singapore.
[1990a] "Algebraic Theory of Wavelets. I: Operational Calculus and Complex Structure," Technical Repors Series #17, Univ. of Lowell.
[1990b] "Generalized Wavelet Transforms. I: The Windowed X–Ray Transform," Technical Repors Series #18, Univ. of Lowell.
[1990c] "Generalized Wavelet Transforms. II: The Multivariate Analytic–Signal Transform," Technical Repors Series #19, Univ. of Lowell.
Kirillov, A. A.
[1976] *Elements of the Theory of Representatins,* Springer–Verlag, Heidelberg.
Klauder, J. R.
[1960] Ann. Phys. **11**, 123.
[1963a] J Math. Phys. **4**, 1055.
[1963b] J Math. Phys. **4**, 1058.
Klauder, J. R. and Skagerstam, B.–S., eds.
[1985] *Coherent States,* World Scientific, Singapore.

Klauder, J. R. and Sudarshan, E. C. G.
[1968] *Fundamentals of Quantum Optics,* Benjamin, New York.
Kobayashi, S. and Nomizu, K.
[1963] *Foundations of Differential Geometry,* vol. I, Interscience, New York.
[1969] *Foundations of Differential Geometry,* vol. II, Interscience, New York.
Kostant, B.
[1970] "Quantization and Unitary Representatins," in *Lectures in Modern Analysis and Applications,* R. M. Dudley et al., eds. Lecture Notes in Math. **170**, Springer, Heidelberg.
Lemarié, P. G. and Meyer, Y.
[1986] Rev. Math. Iberoamericana **2**, 1.
Lurçat, F.
[1964] Physics **1**, 95.
Mackey, G.
[1955] *The Theory of Group–Representations,* Lecture Notes, Univ. of Chicago.
[1968] *Induced Representations of Groups and Quantum Mechanics,* Benjamin, New York.
Mallat, S.
[1987] "Multiresolution approximation and wavelets," preprint, GRASP Lab., Dept. of Computer and Information Science, Univ. of Pennsylvania, to be published in Trans. Amer. Math. Soc.
Marsden, J. E.
[1981] *Lectures on Geometric Methods in Mathematical Physics,* Notes from a Conference held at Univ. of Lowell, CBMS–NSF Regional Conference Series in Applied Mathematics vol. **37**, SIAM, Philadelphia, PA.
Meschkowski, H.
[1962] *Hilbertsche Räume mit Kernfunktion,* Springer, Heidelberg.
Messiah, A.
[1963] *Quantum Mechanics,* vol. II, North–Holland, Amsterdam.
Meyer, Y.
[1985] "Principe d'incertitude, bases hilbertiennes et algèbres d'opérateurs," Séminaire Bourbaki **662**.
[1986] "Ondelettes et functions splines," Séminaire EDP, Ecole Polytechnique, Paris, France (December issue).
Misner, C., Thorne, K. S. and Wheeler, J. A.
[1970] *Gravitation,* W. H. Freeman, San Francisco.

Nelson, E.
[1959] Ann. Math. **70**, 572.
[1973a] J. Funct. Anal. **12**, 97.
[1973b] J. Funct. Anal. **12**, 211.
Neumann, J. von
[1931] Math. Ann. **104**, 570.
Newton, T. D. and Wigner, E.
[1949] Rev. Mod. Phys. **21**, 400.
Onofri, E.
[1975] J. Math. Phys. **16**, 1087.
Papoulis, A.
[1962] *The Fourier Integral and its Applications,* McGraw–Hill.
Perelomov, A. M.
[1972] Commun. Math. Phys. **26**, 222.
[1986] *Generalized Coherent States and Their Applications,* Springer, Heidelberg.
Prugovečki, E.
[1976] J. Phys. **A9**, 1851.
[1978] J. Math. Phys. **19**, 2260.
[1984] *Stochastic Quantum Mechanics and Quantum Spacetime,* D. Reidel, Dordrecht.
Robinson, D. W.
[1962] Helv. Physica Acta **35**, 403.
Roederer, J. G.
[1975] *Introduction to the Physics and Psychophysics of Music,* Springer, New York.
Schrödinger, E.
[1926] Naturwissenschaften **14**, 664.
Segal, I. E.
[1956a] Trans. Amer. Math. Soc. **81**, 106.
[1956b] Ann. Math. **63**, 160.
[1958] Trans. Amer. Math. Soc. **88**, 12.
[1963a] Illinois J. Math. **6**, 500.
[1963b] *Mathematical Problems of Relativistic Physics,* Amer. Math. Soc., Providence.
[1965] Bull. Amer. Math. Soc. **71**, 419.
Simms, D. J. and Woodhouse, N. M. J.
[1976] *Lectures on Geometric Quantization,* Lecture Notes in Physics **53**, Springer Verlag, Heidelberg.

Śniatycki, J.
[1980] *Geometric Quantization and Quantum mechanics,* Springer, Heidelberg.
Souriau, J.–M.
[1970] *Structure des Systèmes Dynamiques,* Dunod, Paris.
Stein, E.
[1970] *Singular Integrals and Differentiability Properties of Functions,* Princeton Univ. Press, Princeton.
Strang, G.
[1989] SIAM Rev. **31**, 614.
Streater, R. F.
[1967] Commun. Math. Phys. **4**, 217.
Streater, R. F. and Wightman, A. S.
[1964] *PCT, Spin & Statistics, and All That,* Benjamin, New York.
Thirring, W.
[1980] *Quantum Mechanics of Large Systems,* Springer, New York.
Toller, M.
[1978] Nuovo Cim. **44B**, 67.
Unterberger, A.
[1988] Commun. in Partial Diff. Eqs. **13**, 847.
Varadarajan, V. S.
[1968] *Geometry of Quantum Theory,* vol. I, D. Van Nostrand, Princeton, NJ.
[1970] *Geometry of Quantum Theory,* vol. II, D. Van Nostrand, Princeton, NJ.
[1974] *Lie Groups, Lie Algebras, and their Representations,* Prentice–Hall, Englewood Cliffs, NJ.
Warner, F.
[1971] *Foundations of Differentiable Manifolds and Lie Groups,* Scott, Foresman and Co., London.
Wells, R. O.
[1980] *Differential Analysis on Complex Manifolds,* Springer, New York.
Weyl, H.
[1919] Ann. d. Physik **59**, 101.
[1931] *The Theory of Groups and Quantum Mechanics,* translated by H. P. Robertson, Dover, New York
Wightman, A. S.
[1962] Rev. Mod. Phys. **34**, 845.
Yosida, G.
[1971] *Functional Analysis,* Springer, Heidelberg.

Young, R. M.
[1980] *An Introduction to Non–Harmonic Fourier Series,* Academic Press, New York.
Zakai, M.
[1960] Information and Control **3**, 101.